

J. Linguistics 42 (2006), 25–70. © 2006 Cambridge University Press
doi:10.1017/S0022226705003683 Printed in the United Kingdom

Against markedness (and what to replace it with)¹

MARTIN HASPELMATH

Max-Planck-Institut für evolutionäre Anthropologie, Leipzig

(Received 18 January 2005; revised 12 May 2005)

This paper first provides an overview of the various senses in which the terms ‘marked’ and ‘unmarked’ have been used in 20th-century linguistics. Twelve different senses, related only by family resemblances, are distinguished, grouped into four larger classes: markedness as complexity, as difficulty, as abnormality, and as a multidimensional correlation. In the second part of the paper, it is argued that the term ‘markedness’ is superfluous, because some of the concepts that it denotes are not helpful, and others are better expressed by more straightforward, less ambiguous terms. In a great many cases, frequency asymmetries can be shown to lead to a direct explanation of observed structural asymmetries, and in other cases additional concrete, substantive factors such as phonetic difficulty and pragmatic inferences can replace reference to an abstract notion of ‘markedness’.

1. INTRODUCTION

The goal of this contribution is to shift the burden of proof onto those linguists who want to make use of a markedness concept in their theoretical work. Given the huge amount of writings about markedness, I cannot possibly aim to argue against every proposal found in the literature, but I will try to identify and present the major markedness notions, refer to the most influential works that have advocated them, and give reasons why I believe that the relevant linguistic phenomena do not require a notion of markedness to understand them. Most of the arguments against markedness come from the earlier literature, so my own contribution here consists primarily in pulling them together and highlighting the radical conclusion that markedness should be abandoned by linguists.

Since it was first proposed by Nicholas Trubetzkoy and Roman Jakobson in the 1930s, the term ‘markedness’ has been very popular in linguistics.

[1] Versions of this paper were presented at the Max Planck Institute for Evolutionary Anthropology, Stanford University, the University of Stockholm, and The Free University of Berlin. I am grateful to the audiences at those occasions. Further useful comments were provided by Juliette Blevins, Grev Corbett, Michael Cysouw, and Anette Rosenbach. I am grateful to everyone who contributed to improving this paper, including two *JL* referees.

MARKEDNESS AS COMPLEXITY

- | | |
|---|---|
| 1. Trubetzkoyan markedness:
markedness as specification
for a phonological
distinction | ‘In German, the phonological
opposition <i>t:d</i> is neutralized syllable-finally
in favor of <i>t</i> , which shows that <i>d</i> is the
mark-bearing member of the opposition.’ |
| 2. Semantic markedness:
markedness as specification
for a semantic distinction | ‘In the English opposition <i>dog/bitch</i> , <i>dog</i> is
the unmarked member because it can refer
to male dogs or to dogs in general.’ |
| 3. Formal markedness:
markedness as overt coding | ‘In English, the past tense is marked
(by <i>-ed</i>) and the present tense is unmarked.’ |

MARKEDNESS AS DIFFICULTY

- | | |
|--|--|
| 4. Phonetic markedness:
markedness as phonetic
difficulty | ‘On the scale $b > d > g > G$, the
consonants to the right are
increasingly more marked.’ |
| 5. Markedness as morphologi-
cal difficulty/unnaturalness | ‘A singular/plural pair like <i>book/books</i> is
less marked than <i>sheep/sheep</i> because
the latter is not iconic.’ |
| 6. Cognitive markedness:
markedness as conceptual
difficulty | ‘The plural category is marked because
it requires more mental effort and
processing time than the singular.’ |

MARKEDNESS AS ABNORMALITY

- | | |
|---|--|
| 7. Textual markedness:
markedness as rarity in texts | ‘For direct objects, coreference with the
subject is marked and disjoint reference is
unmarked.’ |
| 8. Situational markedness:
markedness as rarity
in the world | ‘For marked situations, languages
typically use complex expressions.’ |
| 9. Typological markedness:
markedness as typological
implication or cross-
linguistic rarity | ‘The syllable coda position is marked
in contrast to the onset position.’ |
| 10. Distributional markedness:
markedness as restricted
distribution | ‘Object-verb word order is the marked case:
it occurs only with negation.’ |
| 11. Markedness as deviation
from default parameter
setting | ‘Absence of noun incorporation is the
unmarked case, and the presence of
productive noun incorporation has to
be triggered by a specific parametric
property.’ |
| 12. MARKEDNESS AS A
MULTIDIMENSIONAL
CORRELATION | ‘The singular is more marked than the
plural, and the plural is more marked
than the dual.’ |
-

Table 1

Twelve senses of ‘markedness’ and their typical uses

It was embraced by European structuralism, generative phonology, functional-typological linguistics, Chomskyan principles-and-parameters syntax, neo-Gricean pragmatics, Optimality Theory, first and second language acquisition, creole studies, and probably other research areas as well. In the course of this process, two things happened. First, not surprisingly, the term ‘markedness’ developed a multiplicity of sometimes widely diverging senses that linguists who use it are often not aware of. Secondly, ‘markedness’ lost its association with a particular theoretical approach and became established as an almost theory-neutral everyday term in linguistics. Even though most linguists will be happy to accept that ‘markedness’ has sometimes been used inappropriately, I expect many readers to be surprised that it should be possible to argue for the downright elimination of ‘markedness’ from linguists’ theoretical arsenal.

In section 2 of this paper, I will describe and illustrate twelve different senses of the term ‘markedness’ that are found in the literature. This list should be reasonably exhaustive, although sub-senses could of course be distinguished in a more detailed study. In sections 4–8, I will then explain for each of these senses why in my view they are not required for the linguist’s task of describing and understanding language(s). In each case, I will say what ‘markedness’ should be replaced with – primarily explanations in terms of substantive factors like frequency of use, phonetic difficulty, and generalized conversational implicatures. Table 1 gives a preliminary overview of the twelve markedness senses, with each sense accompanied by a typical specimen of its use in the linguistic literature. (Another, more detailed overview is provided in table 4, at the end of this paper.)

2. TWELVE SENSES OF THE TERM ‘MARKEDNESS’

‘Markedness’ is a polysemous term in linguistics. The various senses with which it is used are connected through their historical origins (ultimately in Trubetzkoy’s and Jakobson’s work of the 1930s) and synchronically through family resemblances. However, most linguists who use the terms ‘marked/unmarked’ use them only in one or a subset of the various senses, and often they do not seem to be aware that the other senses exist, or that the differences between the senses can be rather dramatic (but see Zwicky 1978 and Gair 1988 for earlier lists of different markedness senses). There are very few works that try to work with a concept of markedness which subsumes all or at least a large part of the diverse senses of ‘markedness’ (but see Moravcsik & Wirth 1986, Battistella 1990, Andersen 2001, Ludwig 2001). In this section, I distinguish twelve different senses, grouped into four larger classes: markedness as complexity, as difficulty, as abnormality, and as a multi-dimensional correlation. In the final subsection, I briefly discuss the role of markedness in Optimality Theory.

2.1 *Markedness as complexity*

Under this general heading I discuss three of the twelve senses of markedness, including the sense with which the term was first used.

2.1.1 *Sense 1: markedness as specification for a phonological distinction*

This is the original sense of the term as a theoretical concept, found in Trubetzkoy's work (Trubetzkoy 1931, 1939).² In discussing phoneme inventories, Trubetzkoy distinguishes various types of 'oppositions' (relations within phoneme pairs), among them 'privative oppositions', i.e. oppositions in which one member can be said to bear a 'mark' (German *Merkmal*), while the other member lacks it. Some examples that he gives are in (1) (Trubetzkoy 1939: 67).

(1)	MARK-BEARING	MARKLESS
	(<i>merkmaltragend</i>)	(<i>merkmallos</i>)
	voiced	voiceless
	nasalized	non-nasalized
	rounded	unrounded

Trubetzkoy's notion of 'mark' is an abstract one that cannot be read off directly from a segment's phonetic properties. With regard to the contrast between *t* and *d*, for instance, he notes that *t* is usually tense, while *d* is non-tense (or lax), and that only the 'structure and the functioning of the system' tells us whether it is *d* that is mark-bearing (the mark being 'voice') or whether *t* is mark-bearing (the mark being 'tenseness'; Trubetzkoy 1939: 68). 'Functioning of the system' seems to refer primarily to neutralization (see section 2.3.4): in a language like German, where the syllable-final position only allows *t*, the mark-bearing member of the opposition is *d*.

This sense of 'mark' was not widely adopted by later phonologists, so I will not specifically argue against it here.³

2.1.2 *Sense 2: markedness as specification for a semantic distinction ('semantic markedness')*

Jakobson (1932, 1939, 1957) adopted Trubetzkoy's notion of mark and applied it to oppositions of lexical and grammatical meaning such as those between male and female animal names and between perfective and imperfective

[2] See Gvozdanović (1989) for detailed discussion of Trubetzkoy's markedness concept and its problems.

[3] However, the idea that phonological oppositions (or features) should be characterized by the presence of a specification versus its absence has been influential also in more recent times, especially under the headings of 'radical underspecification' (e.g. Archangeli 1988) and unary features (often called, confusingly, 'privative features'; e.g. Steriade 1995).

aspect in Russian.⁴ Thus, using the example of *osěl* ‘donkey’ and *oslica* ‘female donkey’, Jakobson (1932[1971: 3f.]) said that *oslica* bears a kind of semantic mark indicating female sex, whereas *osěl* lacks any specification for sex. For this reason, *osěl* is not incompatible with female sex and can be used not only for male donkeys, but also for the category of donkeys in general (the situation is similar to the pair *dog/bitch* in English). Thus, the difference between marked and unmarked is not that between A and non-A, but between A and indifference between A and non-A. Similarly, Jakobson (1957[1971: 137]) describes the Russian perfective aspect as ‘concerned with the absolute completion’ of an event, whereas the imperfective aspect is ‘noncommittal with respect to completion or non-completion’ (i.e. unmarked).

This semantic sense of mark/markedness is less abstract than Trubetzkoy’s phonological sense, because it is not just defined in terms of the system, but also in substantive terms: the marked member is semantically more specific than the unmarked member.

I argue (in section 5) that the semantic phenomena discussed under ‘semantic markedness’ should be described with standard semantic concepts like hyponymy and polysemy, and that generalized conversational implicatures and their conventionalization are crucial for understanding the observed asymmetries.

2.1.3 *Sense 3: markedness as overt coding* (‘formal markedness’)

When linguists say that a category X is unmarked, while a category Y is marked, they often simply mean that Y is overtly coded by an affix or auxiliary word, whereas X has no such overt coding (‘zero expression’). For instance, Frajzyngier (2001: 56) says in his grammar of the Chadic language Lele: ‘Whenever a noun has a singular-plural distinction, the singular form is unmarked and the plural form is derived from the singular’. This usage of the term ‘(un)marked’ could be considered as fairly trivial and perhaps even unrelated to the technical term ‘markedness’, but note that the non-technical verb ‘to mark’ (= ‘signal’, ‘code’, ‘indicate’) can also be used when a distinction is made clear by the lack of overt coding: linguists routinely talk about categories being ‘marked by zero’, for instance. Thus, ‘(un)marked’ in the sense of ‘(not) overtly coded’ is different from ‘marked’ in the sense of ‘signaled’, ‘coded’, ‘indicated’.

Markedness as coding complexity has sometimes been understood in a broader sense than simply overt vs. zero expression. Thus, Levinson (2000: 137) states that ‘marked forms ... are more morphologically complex and less lexicalized, more prolix or periphrastic’, and an example he gives is

[4] Jakobson’s markedness concept (or rather, concepts) is discussed in detail by Battistella (1996: ch. 2). A very critical evaluation of Jakobson’s approach is found in Andersen (2001).

The outlaw killed the sheriff vs. *The outlaw caused the sheriff to die* (cf. also Horn 1984: 22). The second, ‘more marked’, construction is not formally related to the first one at all – it only has greater phonetic length.

Although ‘marked’ in the sense ‘overtly coded’ looks like a harmless descriptive term with no theoretical baggage attached to it, it is better to say ‘overtly coded’ instead of ‘marked’ (cf. Croft 2003: 92), and ‘uncoded’ (or ‘zero-coded’) instead of ‘unmarked’, in order to avoid possible misunderstandings.⁵ For the broader sense, ‘phonetically longer’ should be used if this is what is intended.⁶

2.2 *Markedness as difficulty*

Trubetzkoy’s markedness notion was language-particular and purely structural, and substantive considerations were secondary for it. Jakobson (1941, 1963), however, observed that the marked members of oppositions were acquired later by children and were found in fewer languages, suggesting that they are not only more complex in their abstract structure, but also more difficult for language users.

2.2.1 *Sense 4: markedness as phonetic difficulty*

A typical statement is the following:

Markedness: the tendency for phonetic terms to be pronounced in a simple, natural way (as determined in part by the nature of speech articulation, acoustics, and audition, and in part perhaps by more abstract cognitive factors – all aspects of the human language faculty). (Anderson & Lightfoot 2002: 101)

Phonologists have often avoided speaking of plain ‘ease’ or ‘difficulty’ as explanatory factors, and have resorted to terms such as ‘(un)natural’ (as in Natural Phonology, e.g. Hurch & Rhodes 1996), ‘(dis)preferred’ (e.g. Vennemann 1988), or ‘(un)marked’, but it seems that it is precisely ease/difficulty for the language user that is at the heart of these notions (cf. Mayerthaler 1987: 27, who states that more or less natural/unmarked ‘really boils down to more or less easy for the human brain’). Over the last decade, phonologists have often used the term ‘grounding’ when referring to the role of system-external factors in explaining phonological patterns

[5] Faingold (2003: 6) uses the term ‘(un)marked’ for this sense, in order to distinguish it from other senses of ‘(un)marked’. Similarly, Wurzel (1998: 61) proposes the terms ‘featured/featureless’ for senses 1–3, as opposed to ‘(un)marked’ for senses 4–5.

[6] However, Levinson also includes lexical pairs like *book/tome*, *house/residence*, *letter/missive* in his discussion, suggesting that for him the crucial factor is not length, but frequency of use (sense 7).

(e.g. Archangeli & Pulleyblank 1994, Bermúdez-Otero & Börjars 2006). Thus, when Ito & Mester (2003: 16) say that ‘dissimilation of identicals is to a significant extent grounded in markedness’, they seem to understand ‘markedness’ as referring to these physiological, acoustic and perceptual factors, rather than to an aspect of the language system. Other phonologists, by contrast, have explicitly taken the position that while markedness ‘transparently reflects’, ‘is based on’, or ‘is rooted in’ (speakers’ knowledge of) phonetic difficulty, it is still part of the phonological system (e.g. Hayes & Steriade 2004). Unfortunately, the precise relation between phonetic difficulty and phonological knowledge is often left unclear in the phonological literature on markedness.

I argue (in section 6) that ‘markedness as phonetic difficulty’ should be replaced by ‘phonetic difficulty’ and other system-external factors (frequency effects, regularity of sound change).

2.2.2 *Sense 5: markedness as morphological difficulty (dispreference/unnaturalness)*

In the writings of natural morphologists (Mayerthaler 1981, Dressler et al. 1987, Kilani-Schoch 1988, Wurzel 1994, 1998, Faingold 2003), the concept of phonological markedness as ‘unnaturalness’ or difficulty is carried over to morphological structures. The idea is that certain morphological structures are less preferred than others, because they ‘strain the human language capacity’ (Wurzel 1998: 63). In particular, morphological structures are said to be preferred if they have the properties in (i)–(iii).

- (i) CONSTRUCTIONAL ICONICITY, i.e. semantic markedness in the sense of section 2.1.2 (or section 2.2.3) correlates with formal markedness in the sense of section 2.1.3. For example, English singular–plural pairs like *girl/girl-s* are iconic and hence maximally unmarked (= natural), pairs like *sheep/sheep* are non-iconic and hence more marked, and Welsh singular–plural pairs like *pluen* ‘feather’/*plu* ‘feathers’ are counter-iconic and hence highly marked.
- (ii) UNIFORMITY, i.e. morphemes have invariant shapes. For example, German singular–plural pairs like *Boot/Boote* ‘boat(s)’ are less marked than pairs like *Zug/Züg-e* ‘train(s)’, where the plural stem is different.
- (iii) TRANSPARENCY, i.e. an exponent has only one function. For example, German verb inflection is less marked in the singular with transparent coding (*leb-e/leb-st/leb-t* ‘(I) live/(thou) livest/(she)lives’) than in the plural, where the 1st and 3rd persons are syncretic and thus not fully transparent (*leb-en/leb-t/leb-en* ‘(we) live/(you) live/(they) live’).

The markedness or difficulty of structures that violate iconicity, uniformity and transparency is claimed to be reflected in a variety of different

domains: unmarked morphological structures are claimed to (i) be widely found cross-linguistically, (ii) be acquired early, (iii) be processed more easily, (iv) be affected less by language disorders, (v) be used more frequently, and (vi) be more resistant to language change (cf. Mayerthaler 1981: 4f., Faingold 2003: 3–8).

I argue (in section 7) that the ‘unnatural’ phenomena discussed under the rubric of markedness can all be explained with reference to concrete substantive factors like token frequency, type frequency, and regularities of language change.

2.2.3 *Sense 6: markedness as conceptual difficulty*

According to Givón (1991: 337), ‘[t]he marked category tends to be cognitively more complex – in terms of attention, mental effort or processing time – than the unmarked one’. For instance, Givón claims that passive structures are more difficult to process than active structures (p. 352), that the instrument role exhibits lower ‘cognitive accessibility’ than the agent role (p. 354), that definite NPs are ‘cognitively more complex’ than indefinite NPs (p. 355), and so on. Similarly, Langacker (1991: 74) says about the singular–plural distinction:

It is natural that ... it should be the singular that is left unmarked ... the conception of a single instance is simpler than one encompassing multiple instances.

and R. T. Lakoff (2000: 44) says about the markedness of the English past-tense suffix *-ed*:

This extra morphology makes sense to us as speakers of English, because it seems intuitive to us that ‘now’ is simpler than ‘then’. It is, after all, what we are experiencing as we speak, not something that we have to strain to remember or imagine.

Some of these quotations contain the words ‘simple’/‘complex’ rather than ‘easy’/‘difficult’, but these authors clearly do not intend semantic-conceptual markedness as merely consisting in additional specification or narrowing, as in section 2.1.2.⁷

[7] A somewhat different attempt at unifying the semantic relations between grammatical categories (like present/past, imperfective/perfective, nominative/ergative, animate/inanimate, first/second, subject/object, definite/indefinite, singular/plural) was made by Bechert 1977[1998] and adopted by Mayerthaler (1981: 11–21) (see also Wurzel 1994: 2592): The idea is that the unmarked categories in all cases reflect the prototypical characteristics of the speaker, who is animate, first person, singular, etc. ‘and perceives some phenomena more easily than others’ (Wurzel 1994: 2592). It seems that ultimately this approach, too, relates semantic markedness to conceptual difficulty.

I argue (in section 4.4) that ‘conceptual difficulty’ may be caused by low frequency of use, and that it is often the cause for it. There is no need for a ‘markedness’ concept to mediate between cause and effect.

2.3 *Markedness as abnormality*

Complexity and difficulty of linguistic structures has been found to correlate with ‘abnormality’ in various senses, and ‘marked’ therefore has come to mean ‘abnormal’ as well.

2.3.1 *Sense 7: markedness as rarity in texts*

This is the easiest kind of abnormality to observe. Already Trubetzkoy (1939: 235f.) was aware that there are systematic skewings in the text frequency of different phonemes, and he observed that mark-bearing phonemes (see section 2.1.1 above) tend to be rarer than markless phonemes. Especially Greenberg (1966) emphasized the importance of frequency for markedness asymmetries, and he was the first to assign it an explanatory role in this context. Text frequency/rarity has sometimes been taken as the basis for definitions of markedness, although for some reason many linguists seem to prefer vaguer terms like ‘typical’, ‘normal’, ‘usual’ to the precise term ‘frequent’:

To some extent, we can equate the term ‘unmarked’ with ‘regular’, ‘normal’, ‘usual’; and ‘marked’ with ‘irregular’, ‘abnormal’, ‘exceptional’, or ‘unusual’. (Radford 1988: 39)

[T]he typical pattern or property is called UNMARKED, the atypical one MARKED. (Archangeli 1992: 391)

Another example is Baayen et al. (1997: 14), who explicitly define ‘marked form’ (of a singular-plural pair) as the form which occurs less frequently.⁸

I argue (in section 4), following Zipf, Greenberg, and others, that frequency of use is an extremely important factor for explaining a wide variety of asymmetries in language structure. There is no reason why we should not use the words ‘frequent’ and ‘rare’ when we intend them.

[8] Confusingly, some authors say ‘semantic/conceptual/pragmatic markedness’ when they are in fact referring to rarity in texts (of semantic/conceptual/pragmatic categories). For instance, Haiman (1985: 148) rejects the identification of semantic markedness with semantic complexity: ‘a concept may be marked because it is relatively unfamiliar or infrequent’ (his example is *female hippo*, which is not semantically more complex than *mare*). And Bybee (1985: 147) says that for stative verbs, the imperfective aspect ‘is the conceptually unmarked member’, because they occur more often in the imperfective aspect than in the perfective aspect (for reasons having to do with the conceptual content of these categories).

2.3.2 *Sense 8: markedness as rarity in the world*

Some linguists extend this use of ‘markedness’ to extralinguistic rarity. For example, Comrie (1986: 87) observes for Armenian ‘a correlation between the formal markedness of the locative construction and the degree of markedness of the locational situation in the world being described’. Armenian can use the simple locative case for ‘in the box’ (*tup^h-um* ‘box-LOC’), but must use a (formally more marked) postposition for a ‘less natural’⁹ locational orientation such as ‘on the box’ (*tup^h-i vəra* ‘box-GEN on’).

Levinson (2000: 136) uses the expression ‘marked situation’ in the formulation of one of his central principles, as a synonym of ‘abnormal’ or ‘rare in the world’:¹⁰

What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations.

I argue in section 4.2 that rarity/frequency in the world is strictly speaking irrelevant for linguistics. All that matters is text frequency, which is sometimes, but by no means always, correlated with frequency in the world.

2.3.3 *Sense 9: markedness as typological implication or cross-linguistic rarity* (‘*typological markedness*’)

Since Jakobson (1941), phonologists have been aware that the existence of more complex or more difficult phonemes in a language generally implies that corresponding simpler or easier phonemes exist as well. According to Kean (1992: 390),

perhaps the most common view of markedness encountered in the literature is the one based on cross-linguistic distributional analysis ... [e.g.] if a language has a voiced stop, then it has a voiceless one as well.

Similarly, Archangeli (1997: 2) says at the very beginning of a textbook:

The term markedness is used to refer to [the continuum between language-universal and language-particular properties], with completely unmarked properties being those found in virtually all languages and extremely marked properties found quite rarely.

[9] Note that ‘natural’ is used as a quasi-synonym of ‘frequent’ here, i.e. in a rather different sense from that found in section 2.2.1 and section 2.2.2. (‘Naturalness’ is of course another term that should be avoided by linguists.)

[10] When ‘markedness’ is applied to kinds of language use, we have a sense intermediate between 7 and 8. For example, Givón (1991: 335) says: ‘Formal academic discourse is a *marked* discourse type in comparison with everyday oral communication’. Similarly, Waugh (1982: 308) says that “‘spoken” language is the unmarked term and “written” language is the marked term’. (See also Andersen 2001: 24–28.)

And Kager (1999: 11) says that markedness constraints in Optimality Theory may be ‘validated’ by cross-linguistic studies, in combination with phonetic grounding (as mentioned in section 2.2.1). The use of ‘markedness’ in generative phonology ultimately goes back to Chomsky & Halle (1968: ch. 9), where markedness values were a technical device to capture the relative ‘naturalness’ of phonological structures, and ‘naturalness’ is apparently a synonym for cross-linguistic frequency or expectedness.¹¹

I argue (in section 6) that typological implication is merely a consequence of phonetic difficulty or other substantive factors.

2.3.4 *Sense 10: markedness as restricted distribution* (‘distributional markedness’)

More complex and more difficult segments are often restricted in their phonotactic distribution. For example, palatalized obstruents do not occur before liquids in Russian. In German, voiced obstruents may only occur in the syllable onset and are banned from the coda. An example from lexical semantics is provided by Lyons (1977: 306f.): English *dog* shows a wider distribution than *bitch* in that it can be combined with the adjectives *male* and *female* (*male dog*, *female dog* vs. **male bitch*, **female bitch*). More interestingly, in gradable antonym pairs like *high/low*, *old/young*, only one member normally occurs in degree questions like *How old is she?*

The positions where only one member of a pair can occur are said to exhibit ‘neutralization’ of the opposition, and this was Trubetzkoy’s main criterion for assigning phonological markedness values (as we saw in section 2.1.1). Thus, restricted distribution has been important in determining markedness from the very beginning, but it has been taken as the sole definitional criterion only with respect to syntactic constructions. Thus, Dixon (1994: 56ff.) defines ‘functional markedness’ (which he opposes to ‘formal markedness’, i.e. our sense 3) as restricted distribution.

There are two different situations in which two competing syntactic constructions A and B have been said to stand in a distributional markedness relation. On the one hand, if B only occurs under specified conditions, while A may always occur, B is said to be marked and A unmarked. An example is German object–verb order in subordinate clauses: the order verb–object

[11] This sense of markedness has also been influential in the literature on second language acquisition. According to Eckman (1977: 320), ‘A is typologically marked relative to B iff every language that has A also has B but not every language that has B also has A’, and based on this, he proposes the Markedness Differential Hypothesis, which essentially says that target language structures which are more marked than native language structures will be difficult for second language learners, while target language structures which are not more marked will not be difficult (see Rutherford 1982, Eckman 1996, 2004 for further discussion). Typological markedness is thus used as a predictor of markedness as difficulty here.

occurs only if the object is heavy, while the order object–verb may always occur, so the latter order is unmarked (Lenerz 1977: 27, see also Müller 1999: 782). On the other hand, a markedness relation has also been diagnosed when B occurs under specified conditions, and A occurs elsewhere. For instance, in the Mayan language Mam, the antipassive construction is used when the object is not mentioned, when the agent is focused, and in a few other cases, while the ergative construction is used elsewhere, so the ergative construction is regarded as unmarked (Lazard 1989: 321). This definition of ‘unmarked’ as ‘the elsewhere case’ is also found in Dryer’s (1995) detailed discussion of markedness in word order constructions. Strictly speaking, on the latter definition, BOTH competing constructions are distributionally restricted, but the ‘unmarked’ construction is restricted negatively by the marked construction, not by any positive specification (thus, ‘marked’ in the sense of ‘positively specified’ is similar to the markedness senses 1 and 2).

‘Marked’/‘unmarked’ in this distributional sense is a relatively harmless term that does not suggest a high degree of abstraction or explanatory depth. Nevertheless, to avoid misunderstandings, it is better to say ‘distributionally unrestricted’ or ‘occurs elsewhere’, rather than ‘unmarked’.

2.3.5 *Sense 11: markedness as deviation from default parameter setting*

In the Chomskyan principles-and-parameters approach, a completely novel use of ‘markedness’ appeared: the idea that the possible parameter settings are not equal, but that (assuming that the choice is binary) one parameter value (the unmarked one) is chosen by default, while the other value (the marked one) is chosen only if the evidence forces the choice.¹²

According to Chomsky (1981: 8), the theory of markedness ‘imposes a preference structure on the parameters of UG [Universal Grammar] ... In the absence of evidence to the contrary, unmarked options are selected’. In other words, ‘the unmarked case of any parameter represents the initial hypothesis that children make about the language to be acquired’ (Kean 1992; see also Haider 1993: 635).

Unmarked parameter values have been said to emerge in creolization: Bickerton’s (1984) bioprogram has been interpreted as the set of unmarked

[12] In Chomsky’s work of the late 1970s and early 1980s, yet another sense of ‘markedness’ is found, in the context of the distinction between core grammar (determined by UG) and a ‘marked periphery’ of rules that are outside the scope of UG. This type of markedness also represents a kind of abnormality, but it was always unclear how core phenomena could be distinguished from peripheral phenomena (Chomsky 1982: 110; cf. the discussion in Botha 1989: 86f., Battistella 1996: 80–89), so this usage of ‘marked’ and the entire core–periphery distinction was quietly abandoned. (In the 1990s and 2000s, ‘markedness’ ceased to play a prominent role in Chomskyan syntax, and recent textbooks no longer even mention the term, cf. Haegeman 1994, Carnie 2002, Radford 2004.)

parameter settings of UG (Battistella 1996: 103), and other linguists have made similar claims about creoles (e.g. Roberts 1998).

Examples of marked parameter settings are noun incorporation ('absence of noun incorporation is the unmarked case, and the presence of productive noun incorporation has to be triggered by a specific parametric property', Baker 1996: 283), preposition stranding (van Riemsdijk 1978), and extraction from finite embedded clauses (Haider 1993: 643).

This model has also been adopted in phonology. For example, Blevins (1995: 218) discusses syllable structure typology in terms of six basic parameters, four of which are illustrated in the partial table in (2).

(2)	COMPLEX NUCLEUS	OBLIGATORY ONSET	COMPLEX ONSET	CODA
Totonac	Y	Y	Y	Y
Klamath	Y	Y	no	no
English	Y	no	Y	Y
Cayuvava	no	no	no	no

Blevins comments: 'The parameter settings described above not only account for the generalizations noted [earlier], but are also meant to encode markedness values, where 'no' is the unmarked value and 'yes' is the marked value'.

A very similar position is expressed by Jackendoff (2002: 191):

[T]he 'unmarked' case is the one prespecified by UG, and 'marked' rules deviate from the unmarked case qualitatively in just the same way as irregular verbs deviate from regular forms ... UG renders certain parts of the design space for words and l-rules more stable and/or accessible, with gradients of relative 'markedness' as one moves away from the core cases.

The evidence in favor of these various proposals is almost always extremely indirect and thus very difficult to evaluate. Sometimes (e.g. in the case of noun incorporation) the main argument for marked status appears to be its cross-linguistic rarity. In such cases, this kind of markedness is a variant of sense 9 (typological implication) with additional assumptions about the source of the asymmetries. But markedness as non-default parameter setting is very different from all other markedness senses because its domain is not linguistic categories, but cognitive states. It seems that 'markedness' in this sense is best understood as a metaphor derived from other senses of 'markedness', in particular the 'non-default' sense of distributional markedness. I will not discuss this sense of 'markedness' any further here.

2.4 *Sense 12: markedness as a multidimensional correlation*

While the various senses of markedness in sections 2.1–2.3 are clearly distinguishable and to a large extent logically independent of each other, they are mostly not incompatible with each other. Thus, markedness could be

defined in a very broad way, as the conjunction of (several or all of) these senses. This is the approach of Greenberg (1966), followed especially by Croft (1990, 2003). But the combination of the various markedness dimensions is not just definitional: according to the multidimensional view of markedness, it is a remarkable observation that universally, comparable linguistic structures exhibit the same markedness values for the different markedness dimensions (or ‘criteria’). Thus, some categories (such as the plural or the future tense) are semantically complex, overtly coded, rare in texts, found only in some languages, and restricted in their distribution (i.e. marked in all these respects), whereas other categories (e.g. the singular or the present tense) are semantically simple, not overtly coded, frequent in texts, found in all or most languages, and unrestricted in their distribution. This is not logically necessary, so it is an important empirical discovery. The following is a list of the properties that Greenberg (1966) discussed; occasionally Croft’s terminology and definitions are used. Not all of these properties were discussed earlier.

- (i) **TEXT FREQUENCY.** This corresponds to sense 7 above: ‘If tokens of a typologically marked value of a category occur at a certain frequency in a given text sample, then tokens of the unmarked value will occur at least as frequently in the text sample’ (Croft 2003: 110).
- (ii) **STRUCTURAL CODING.** This corresponds to sense 3 (section 2.1.3): ‘The marked value of a grammatical category will be expressed by at least as many morphemes as is the unmarked value of that category’ (Croft 2003: 92).
- (iii) **INFLECTIONAL DIFFERENTIATION.** Croft (2003: 97) subsumes Greenberg’s three markedness characteristics of syncretism, defectiveness, and allomorphy under this category: ‘If the marked value has a certain number of formal distinctions in an inflectional paradigm, then the unmarked value will have at least as many formal distinctions in the same paradigm’ (Croft 2003: 97; the original generalization goes back to Brøndal 1940). For instance, German articles make a three-way gender distinction in the singular, but **SYNCRETIZE** the gender distinction in the (marked) plural. French verbs have a special subjunctive mood in the present tense and past tense, but the (marked) future tense is **DEFECTIVE** in that it has only an indicative mood form. In Sanskrit, the (marked) dual case-endings show much less **ALLOMORPHY** than the plural endings. This markedness correlate was not mentioned earlier, because markedness is never equated with reduced inflectional differentiation. But for Greenberg, this criterion was very important.¹³

[13] Greenberg also mentions two other characteristics which are of very restricted applicability. One is ‘dominance’, i.e. the use of a single unmarked word in the plural or dual for a

- (iv) **FACULTATIVE EXPRESSION.** This corresponds to sense 2 (section 2.1.2), and it means that the marked category is optional, while the unmarked category occurs both with a general sense, and ('par excellence') in a specific sense opposite to the marked sense. Greenberg gives the example of the optional Korean plural *-tul*. Korean nouns lacking this suffix can be vague with respect to the singular-plural distinction, or they can be interpreted as specifically singular (Greenberg 1966: 28).
- (v) **CONTEXTUAL NEUTRALIZATION.** This corresponds to sense 10 above (section 2.3.4). 'In certain environments the opposition between two or more categories is suppressed, and it is the unmarked member which appears' (Greenberg 1966: 29). Greenberg gives the examples of number neutralization of nouns with numerals in Turkish (where the singular must be used regardless of the number meaning), of number neutralization in Arabic verbal inflection in verb-initial clauses when a nominal subject is present (where the singular must be used regardless of the number of the subject), and of the use of the third person singular in verbs used without a subject ('impersonally') in Latin (Greenberg 1966: 29, 36, 44).
- (vi) **TYPOLICAL IMPLICATION.** This criterion, which corresponds to sense 9 above (section 2.3.3), is not prominent in Greenberg's (1966) discussion. It appears that it is useful primarily for phonology, where universals such as 'if a language has ejective stops, it has plain stops' make meaningful claims. Croft (1990: 83) cites Greenberg's (1963: 94) Universal 94 ('If a language has the category of dual, then it also has the category of plural') as evidence for the markedness of the dual over the plural. But with most pairs of meaningful categories, such implicational claims make little sense. For instance, we cannot say that 'if a language has a future tense, it has a present tense', because a language with a future tense that lacks a present tense is practically impossible by definition. All languages allow the verbal description of ongoing events, and whatever verb form is used to express such events has to qualify as a present tense in a cross-linguistic perspective.¹⁴ Croft (2003: 88) explicitly claims that markedness is predicated of universal conceptual categories, not particular linguistic categories, so that the question whether a language possesses or lacks a category does not arise (he

heterogeneous group or pair (e.g. Spanish *padres* 'fathers' for 'parents'). This is relevant only for a small number of lexical pairs (see the discussion in section 5). The other is 'agreement a potiori', i.e. the use of an 'unmarked' gender in case of a gender conflict (as in Spanish *el hijo y la hija son buenos* 'the son.M.SG and the daughter.F.SG are good.M.PL'; see also Schane 1970). Corbett (1983) discusses resolution rules for gender agreement in some detail (see also Corbett 1991: 290–293) and concludes that markedness is not relevant.

[14] Even in the case of the dual and the plural, it is questionable that languages having a dual but lacking a plural would be logically possible (cf. Dryer 2004).

therefore no longer includes typological implication among his markedness criteria).

While for Greenberg (and Croft) all these correlating dimensions are about equally important, there is one pair of dimensions whose correlation has been particularly highlighted by other authors: that between semantic/conceptual specificity (i.e. my sense 2, section 2.1.2) or semantic/conceptual complexity (i.e. my sense 6, section 2.2.3 above)¹⁵ and overt coding. This correlation is often described with the term ‘iconicity’ (and this is sometimes regarded as an explanation):

From such a tendency towards iconicity in morphology we would expect zero expression of semantically unmarked categories and overt, or non-zero, expression of semantically marked categories. (Zwicky 1978: 137)

If a semantically more marked category is encoded as formally more marked than a less marked category, the encoding is said to be iconic. (Mayerthaler 1987: 48f.)

[The meta-iconic markedness principle:] Categories that are *cognitively* marked tend also to be *structurally* marked. (Givón 1991: 106/1995: 58)

[Iconicity of complexity:] Marked forms and structures are typically both structurally more complex (or at least longer) and semantically more complex than unmarked ones. (Newmeyer 1992: 763)

Iconicity favors the morphological marking of syntactically marked configurations. (Aissen 2003: 449)

I argue (in section 4.5), following Greenberg (1966) and Croft (2003), that the correlating properties of meaningful categories can all be explained on the basis of frequency asymmetries. The first correlating dimension (frequency) explains all the others, and hence there is no need for a term that describes the correlation itself. Iconicity explanations are not necessary either (see also Haspelmath, in preparation).

2.5 *Markedness in Optimality Theory*

With the advent of Optimality Theory (OT) in the mid-1990s, the term ‘markedness’ has become much more prominent in generative linguistics. It most often occurs in the compound ‘markedness constraint’, defined as ‘any constraint that assigns violation-marks to a candidate based solely on its

[15] Recall that Greenberg referred to the semantic criterion as ‘facultative expression’, i.e. he did not think of the marked term as more complex, and he did not see any iconicity here. Croft (2003) does not recognize facultative expression or semantic specificity as a criterion of markedness at all. (And for him, zero expression represents a violation of iconicity, i.e. strict form-function isomorphism.)

output structure, without regard to its similarity to the input' (McCarthy 2002: 14). This is in contrast primarily to faithfulness constraints, which evaluate candidates by comparing them to the input. This notion of 'markedness' is thus a purely formal one, empty of any substance, reduced to the disfavoring of candidates based only on their properties. Of course, in practice constraints with the right substance will work best, so OT practitioners employ constraints like *ONSET* (a syllable must have an onset) and *NoCODA* (a syllable must not have a coda), rather than constraints like 'NoONSET' or 'CODA' (although no higher principle excludes these constraints).

When one studies the most successful markedness constraints that have been proposed, one finds that indeed their substantive content corresponds to kinds of markedness that we saw earlier. In phonology, most of the widely used markedness constraints seem to correspond to (or be 'grounded' in) phonetic difficulty (see section 2.2.1 above), e.g. constraints like *NoCODA*, **NÇ* (a nasal stop must not be followed by a voiceless obstruent), **V_{ORAL}N* (vowels must not be oral before nasal). And at least some of the markedness constraints proposed in morphosyntax concern meaningful categories. Thus, Woolford (2001: 513) has constraints like **ACCUSATIVE* and **ERGATIVE*, Grimshaw (2001: 234) has constraints like **1ST/2ND*, and Aissen (1999) has constraints like **SUBJECT/PATIENT*. These markedness constraints correspond to conceptual difficulty (see section 2.2.3 above) and to the Greenbergian markedness patterns discussed in section 2.4.

Thus, OT markedness constraints largely continue the tradition of the earlier markedness literature. The innovation of OT is restricted to the formal apparatus of constraint interaction. If we want to know why the constraints are the way they are, we are thrown back to the same issues that the earlier literature grappled with: substantive factors like ease/difficulty of perception and production, and frequency of use.

3. SIX ROLES OF MARKEDNESS

Before moving on to show why 'markedness' can be dispensed with, let us consider the various levels at which the term has played a role in linguists' writings. Unfortunately, linguists are often inexplicit about the precise scope of their claims, but the following seems to be a reasonably complete list of the possibilities.

- (i) **MARKEDNESS IS REPRESENTED MENTALLY IN PARTICULAR GRAMMARS.** In Trubetzkoy's and Jakobson's earlier writings, it seems that markedness is conceived of as a language-particular phenomenon (cf. Croft 1996). For instance, in an opposition *t:d*, some languages could have *d* and others could have *t* as the marked member. Similarly, Jakobson's claim about the perfective-imperfective opposition was only about Russian.

Although the Prague School linguists did not emphasize this, their claims seem to have been about speakers' mental grammars.

- (ii) **MARKEDNESS IS AVAILABLE IN THE COGNITIVE CODE (=UNIVERSAL GRAMMAR).** In Chomsky & Halle (1968), the idea was proposed that markedness values are not just present in language-particular mental grammars, but are in some way defined at the level of the innate cognitive code for language ('Universal Grammar'). This position lives on in Optimality Theory in the widespread claim that markedness constraints (as well as the other constraints) are innate and part of UG. And of course markedness in the sense of deviation from default parameter setting (section 2.3.5) is part of the cognitive code.¹⁶
- (iii) **MARKEDNESS IS A METAGRAMMATICAL CONCEPT USED BY LINGUISTS FOR THEIR CONVENIENCE.** This seems to have been Greenberg's (1966) position. He observed that a large number of similar properties of pairs of phonological, grammatical and lexical categories can be subsumed under generalizations formulated in terms of markedness, but he made no specific claims about mental grammars or the cognitive code. Nor did he presuppose that these generalizations must have a uniform explanation – in fact, he explicitly suggested that they do not. Also, markedness as defined by distributional restrictions (especially with respect to competing syntactic constructions, section 2.3.4) is a metagrammatical concept that seems to make no claims about speakers' mental grammars.
- (iv) **MARKEDNESS AS AN EXPLANATORY CONCEPT.** In Wurzel's (1998) view (and that of other linguists working in the Natural Morphology tradition), markedness theory is an explanatory theory consisting of a set of statements about structural preferences in languages. It is itself in need of being explained by neighboring disciplines, so it is a core part of linguistic theory, but it is completely separate from UG, and markedness is not relevant to particular grammars. This view of markedness is thus similar to Greenberg's, but more technical, and it is explicitly claimed that markedness is part of a theory.
- (v) **MARKEDNESS AS A GENERAL PROPERTY OF HUMAN CULTURAL MANIFESTATIONS.** Moravcsik & Wirth (1986) claim that the correlation between the three main markedness properties (familiarity, diversity, and simplicity) is found throughout human culture, e.g. in food: everyday food is more frequent, comes in more different kinds, and is simpler than holiday food. Markedness in language is just an instantiation of

[16] However, according to Wexler & Manzini's (1987) influential Subset Principle, the markedness ordering of parameter settings arises from the learning procedure, and is not directly part of UG.

markedness in this more general sense. The idea that markedness is a highly general property of human culture is already found in Jakobson's famous letter to Trubetzkoy from 1930 (published in Trubetzkoy (1975: 162f.), and cited in Waugh (1982: 300f.), Andersen (1989: 22), Battistella (1990: 5); see also Andersen 2001: 25–28). The terms 'marked' and 'unmarked' have even occasionally been adopted by scholars in other fields such as sociology (e.g. Brekhus 1998).

- (vi) **MARKEDNESS IN A NON-TECHNICAL SENSE.** When linguists say that certain constructions are used 'only under very marked circumstances' (Corston-Oliver 2003: 285), that phonological similarity implies semantic similarity 'in the unmarked case' (Johns 1992: 84), or even that certain sentences are judged by speakers to be 'marked' (e.g. Müller 1999: 782), it seems that they simply mean '(un)usual' in a non-technical sense.

Many readers will probably agree immediately that this last use of '(un)-marked' is pointless and should be avoided, so I will not say more about this. As for the other uses, it is quite impossible to argue conclusively against all of them. Claims about mental grammars, (i), and the cognitive code, (ii), are notoriously hard to falsify, and the superfluousness of markedness as a metagrammatical or explanatory concept, (iii, iv), can be demonstrated only by a comprehensive alternative account of all relevant phenomena. Here, my goal is mainly to sow doubt in the reader's mind and to point to places in the literature where such alternative accounts have been developed.

4. THE EXPLANATORY POWER OF FREQUENCY OF USE

4.1 *From performance to competence*

I will start with frequency of use (first with regard to meaningful categories) because this variable can be shown to make five of the markedness senses of section 2 superfluous. Frequency of use is a property of parole or performance, not of language structure or competence, and throughout the 20th century most linguists have shown little interest in explaining structure in terms of use. But the evidence that frequency is not just one correlate of markedness, but in fact the major determinant of markedness effects in morphosyntax, is by now overwhelming (cf. Werner 1989 and Fenk-Oczlon 1991, who argue specifically against Natural Morphology).

At least since Horn (1921) and Zipf (1935), it has been known (to those who cared) that there is a strong inverse correlation between signal length and signal frequency, both in natural languages and in other human semiotic systems. It has long been clear that this must be due to a rational principle of least effort or economy (see Haiman 1983, Hawkins 2004: chs. 3–4 for more recent discussion).

Two mechanisms for creating such human semiotic systems have also long been recognized: shortening of signs when their frequency (and hence

expectedness) increases, and the creation of more rarely used signs by compounding other signs. For example, the contrast between *egg* (= 'chicken egg') and *ostrich egg* may have come about in two ways: either *chicken egg* was shortened to *egg* when these became much more common (after domesticated chickens were introduced) or the term *ostrich egg* was created by compounding *ostrich* and *egg* used in an extended sense (*egg* may have referred exclusively to chicken eggs at some point). Whichever of these historical scenarios is correct for this particular example, we have a fully satisfactory model of explanation that makes no reference to markedness and whose basic structure can be extended to all other lexical and morpho-syntactic phenomena with similar properties.

That much of morphosyntactic and lexical 'markedness' can be explained by frequency of use was in fact recognized by Greenberg (1966: 65–69), who noted that frequency is 'an ever present and powerful factor in the evolution of grammatical categories and thus helps in explaining the types of synchronic states actually found' (this is elaborated by Croft 1990: 156–160). Thus, I have no substantive disagreement with Greenberg, but I agree with Mańczak's (1970) early criticism that since frequency of use seems to explain most of the observed phenomena, we do not need a 'markedness' concept to understand them.

A phenomenon that is immediately predicted by frequency of use is 'markedness reversal' (e.g. Mayerthaler 1981: 48ff., Tiersma 1982, Croft 1990: 135). Under specific circumstances, we observe unexpected 'marked' behavior of categories that are usually unmarked. For example, in some languages the singular of some nouns is overtly coded and corresponds to an uncoded plural (e.g. Welsh *plu-en* 'feather', *plu-Ø* 'feathers'). Similarly, in some languages the second person in the imperative is uncoded, contrasting with an overtly coded third person (e.g. Latin *lauda-to* 'let her praise', *lauda-Ø* 'praise!'). This situation violates iconicity (see section 2.4), but it is in perfect harmony with the frequency-based explanation: only those nouns that occur more frequently in the plural tend to have uncoded plurals, and only those moods that are used more often in the second person (namely the imperative) have uncoded second-person forms. Even more striking examples come from the lexicon, where we have pairs like *priest/priestess* in male-dominated professions, and *nurse/male nurse* in a female-dominated profession. Witkowski & Brown (1983) show how changes in lexical frequency over time can lead speakers to adopt new forms of lexical expression, adding a modifier as a word becomes less frequent over time or dropping modifiers as it becomes more frequent. Such 'markedness reversals' have been reluctantly admitted by markedness theorists, who recognized that markedness is context-dependent (e.g. Waugh 1982: 307). The only way in which the relevant contexts can be defined is with reference to frequency of use.

The 'alignment of prominence scales' that is highlighted in Aissen's (1999, 2003) work on functionally grounded optimality-theoretic syntax is nothing

more than a formalist expression of the effects of differential frequency of use in different contexts. Aissen states that ‘in many languages, expression of more marked configurations is morphologically more complex than that of less marked ones’ (1999: 674), referring to the tendency for agents to be high on the person, animacy and definiteness scales, and for patients to be low on these scales. But Aissen’s ‘universal theory of markedness’ (1999: 708), in its role of predicting cross-linguistic tendencies, can be replaced by frequency of use and the least-effort principle (agents are simply MORE OFTEN definite, patients are MORE OFTEN inanimate, and so on; see also Haspelmath (2004b) for discussion of Aissen 1999).

4.2 *Accounting for rarity in texts (sense 7) and rarity in the world (sense 8)*

My position here is that frequency/rarity in texts (i.e. primarily in everyday speech) is the basic phenomenon that explains a lot of other linguistic phenomena, whereas frequency/rarity in the world is strictly speaking irrelevant for grammatical theory because frequency in the world does not affect language structure directly (see Werner 1989: 42). Thus I need not account for sense 7 (because I take it as basic), and as a linguist I see no reason to worry much about sense 8.

Of course, from a wider perspective it is an interesting question why certain words and categories are used often, and why others are used rarely, even though we do not need an answer to this question to understand language structure. But the answer will be non-uniform, because speakers’ reasons for choosing words and categories are very heterogeneous. Frequency in the world is of course one relevant factor. The higher textual frequency of the first item in pairs like *cow/bull*, *sheep/ram*, *hen/rooster*, *poet/poetess* of course has to do with the fact that in traditional Western culture, farms have more cows than bulls (Lyons 1977: 308, Fenk-Oczlon 1991: 367), more men than women get recognition for their poetry, and so on. But more often than not, frequency in texts has nothing to do with frequency in the world. For instance, the verb *eat* is much more frequent than *go to the bathroom*, even though the latter activity is presumably just as frequent (Ariel 2004), and *beetle* is much rarer than *dog*, even though the world has many more beetles. Clearly, what we talk about is determined not by the world as such, but by our perception of it and by what we find relevant. The present tense is more frequent than the future tense not because few events occur in the future, but because we cannot talk about most of them. The singular is more frequent than the plural not because the world consists of more individuals than groups, but because humans tend to focus on individuals (and to treat groups as individuals, e.g. *herd*, *battalion*, *cloud*). The numeral *nine* is rarer than *four* because with larger groups, we are less concerned with precise numbers. All of these different explanations are interesting (and they could well be wrong), but they do not matter for explaining grammar.

UNMARKED MEMBER	FREQUENCY	MARKED MEMBER	FREQUENCY	RATIO UNMARKED/ MARKED
<i>long</i>	392	<i>short</i>	198	2.0
<i>high</i>	574	<i>low</i>	286	2.0
<i>wide</i>	165	<i>narrow</i>	54	3.1
<i>strong</i>	197	<i>weak</i>	45	4.4
<i>true</i>	183	<i>false</i>	36	5.1
<i>full</i>	289	<i>empty</i>	54	5.4
<i>beautiful</i>	87	<i>ugly</i>	14	6.2
<i>deep</i>	97	<i>shallow</i>	15	6.5

Table 2

Frequencies of some gradable antonyms in English (from Leech et al. 2001, per million word tokens)

It has sometimes been claimed that frequency of use should not be taken as basic because it is derivative of one of the other manifestations of markedness. Thus, Mayerthaler (1981: 136–140) claims that both unmarkedness as unnaturalness (sense 5) and conceptual unmarkedness (sense 6) explain high frequency in texts, without, however, providing a mechanism. Such a mechanism is provided by Lehrer (1985: 399) (and similarly Waugh 1982: 302): ‘Since the unmarked member may occur in a wider range of contexts and will appear where the contrast is neutralized, it will also be more frequent’. Neutralization may account for a small part of the frequency difference, but a brief look at some figures should convince the reader that it is totally implausible that all of it should be explained in this way. In Greenberg’s (1966: 32) figures, the singular occurs in 70–85% of the cases, the plural in 15–25%. To account for this difference, 40–70% of the contexts would have to be neutralizing. In table 2, we see some figures from Leech et al. (2001) for gradable antonyms in English (the topic of Lehrer’s paper), showing that ‘unmarked’ gradable adjectives are between twice and six times as frequent as their ‘marked’ counterparts. It is clearly impossible to argue that neutralizing contexts like degree questions (*How long is this rope?*) should suffice to account for this striking difference.

Another possibility, which cannot be dismissed so easily, is that shorter expressions occur more frequently simply because they are short, i.e. the direction of causation could in principle be the opposite to that given at the beginning of section 4. In phonology, complexity does seem to be a causal factor for frequency (see section 6 below), but I know of no evidence that it is in morphosyntax. The hypothesis that shortness can lead to frequency can be tested by cross-linguistic frequency counts, in particular by examining frequencies of semantically equivalent expressions that are of different lengths

in the respective languages. It seems that coding length has no or very little influence in cases like the singular-plural distinction, where one could suspect that the English nominal plural is rarer than the singular because of its extra segment (and sometimes extra syllable). However, Greenberg's (1966: 32) data from Sanskrit, Latin and French (where the singular usually has as many syllables and segments as the plural) show that in these languages, too, the plural is much rarer than the singular (see also Haspelmath 1999a and 2004b for similar arguments in syntax).

4.3 *Accounting for overt coding (sense 3)*

The correlation between signal length and signal frequency has already been explained at the beginning of section 4 (see also Haiman 1983, 1985; Croft 2003: 110–117). Zero expression is just a special case of short expression, so frequency explains both the narrower sense (overt coding vs. zero expression) and the broader sense (formal length involving greater processing effort) of formal markedness. As Croft (2003: 115f.) recognizes, frequency correlates not just with the number of morphemes (as he had claimed in Croft 1990: 76), but also with the phonetic length of morphemes. If one wanted to extend the notion of markedness to all situations where one expression is more frequent and hence shorter than another, one would have to say that a markedness relation is even found in pairs like *ten/hundred*, *son/brother*, *stomach/appendix*, *sparrow/albatross*, *comma/apostrophe*, *moon/Uranus*, *walk/gallop*, *flu/leukemia*. Greenberg does so for the first two pairs, but it is clearly more parsimonious to just eliminate 'markedness' and describe and explain these cases directly in terms of frequency and phonetic length.

4.4 *Accounting for conceptual difficulty (sense 6)*

A category that is used more frequently will not just be shorter, but also of course easier to process in a number of ways, than a rarely used category (see e.g. Fenk-Oczlon (1991: 373–381) for the role of frequency in language acquisition, language deficits, processing, and leveling). The connection between frequency of use and ease of processing is well known from the psychological literature. For example, if 'passive clauses are processed with more difficulty and acquired later by children than active clauses' (Givón 1991), this is probably due to their lower frequency, and not to their markedness.

In other cases, 'conceptual difficulty' is apparently the cause for the lower frequency of a category. Langacker (1991: 74) seems to be right in saying that 'the conception of a single instance is simpler than one encompassing multiple instances', or at least this would explain why we use the singular more often than the plural.

Thus, in some cases frequency of use explains conceptual or cognitive ease, in other cases it is the other way round. There is no need for a concept of 'markedness'.

4.5 *Accounting for the multidimensional correlation (sense 12)*

After presenting the markedness correlations, Greenberg (1966: 65) concludes that, in grammar and lexicon, frequency of use is ‘primary’, i.e. explains the other markedness correlates. He even goes so far as to suggest that the terms marked/unmarked should be defined as ‘less frequent/more frequent’ (p. 67). Croft (2003: 102) follows him: ‘The more general theoretical concepts are economic and iconic motivation, not typological markedness’. Mańczak (1970) criticized Greenberg for using the Praguean ‘markedness’ concept for phenomena that can be shown to be derived entirely from frequency, and I believe that Mańczak was essentially right. However, this extension of the term ‘markedness’ had already been begun by Jakobson (1963) for phonology, and Greenberg understandably found the universal markedness predictions of Jakobson’s later work more interesting than the strictly language-particular, purely classificatory markedness conception of the 1930s.

Let us briefly review how frequency of use explains the correlates.

- (i) TEXT FREQUENCY. This is not explained, but is the explanatory factor.
- (ii) STRUCTURAL CODING. As we saw above (section 4.1), if grammars are designed economically, they will use zero-coding for the most frequent category in a set of paradigmatic alternatives. Speakers can (unconsciously) bring this about by shortening earlier overt markers, or by using the most frequent form as the base for new formations.
- (iii) INFLECTIONAL DIFFERENTIATION. The reason why more frequent categories show greater allomorphy or more irregularities is that these are more easily remembered when the category occurs often. By contrast, infrequent forms ‘must follow analogically other parts of the system, while only a fairly frequent form can preserve irregularities’ (Greenberg 1966: 68f.). As Mańczak (1970: 33f.) notes, the same is true in the lexicon: suppletion and other irregularities occur primarily in high-frequency items (cf. *brother/sister, stallion/mare, vs. lion/lioness, male hippo/female hippo*). The same principle applies to syncretism (and defectiveness): syncretism is generally found to a greater extent in rarer inflectional categories (and more generally in rarer words/word types) because it is more difficult to remember the distinct forms when they do not occur often (Croft 2003: 113, Hawkins 2004: ch. 4).

These three properties – frequency, coding and differentiation – are regarded as the core dimensions of markedness by Croft (2003). But one other dimension also has an explanation in terms of frequency, namely facultative expression.

- (iv) FACULTATIVE EXPRESSION. Greenberg gives examples of optional plurals, duals, and future and past tenses. Let us first consider the case of

optional plurals, as in Korean. Here we would probably want to say that the ‘singular’ form (which can also refer to groups) does not in fact have singular meaning, but is a ‘general number’ form (Corbett 2000: 9–19). The reason why it is often interpreted (‘par excellence’) as singular is simply that this is the most frequent situation; for example, when the word *chayk* ‘book’ is used, most of the time a single book is referred to, so speakers adopt this as the default hypothesis and revise this hypothesis only if needed. (See section 5 for further discussion.)

The remaining two dimensions do not seem to have a frequency explanation, but we already saw in section 2.4 that ‘typological implication’ (vi) is not an empirically observable correlate. As for contextual neutralization (cf. (v) in section 2.4), Croft (2003: 100f.) argues that it simply fails to correlate with the other dimensions: ‘there is no cross-linguistic consistency as to which value is chosen’. For instance, in some languages the masculine is used for referents of unknown sex, while in other languages the feminine is chosen (e.g. Jarawara, Dixon 2004: 286). Greenberg himself noted that with respect to syncretism, neuter is the most marked gender in Indo-European, but with respect to contextual neutralization, it patterns as the unmarked gender (1966: 39f.); and in English, the plural pronoun *they* can be used for singular reference when the gender is neutralized.

The only generalization that seems to be fairly robust cross-linguistically is the consistent use of singular number and 3rd person agreement forms when a controller is missing or has no relevant features. In most cases this can be explained by economy, because the 3rd person singular forms usually have the least overt coding, and when no relevant information is conveyed, it would be uneconomical to use overt coding elements when no information is in fact conveyed. However, a small residue of unexplained facts remains, for instance, the fact that in Latin (which has no zero forms in its verbal subject inflection) we find 3rd person, rather than another person, in subjectless verbs like *pluit* ‘it is raining’ or *oportet* ‘it is necessary’. From the point of view of pure economy, one would expect special zero-coded forms here (*plui*, *oportet*).¹⁷ This residue is hardly sufficient to motivate keeping a ‘markedness’ concept in linguistics.

5. REPLACING ‘SEMANTIC MARKEDNESS’

Semantic markedness has sometimes been regarded as an extremely general phenomenon that encompasses much more than just some grammatical and lexical semantic contrasts. Greenberg (1966: 25), for example, invokes

[17] That such forms are impossible seems to be related to the fact that some languages require a dummy subject pronoun, like English *it* or French *il*, in such constructions.

[t]he pervasive nature in human thinking of this tendency to take one of the members of an oppositional category as unmarked so that it represents either the entire category or *par excellence* the opposite member to the marked category

and mentions the role this tendency plays in mathematical symbols ('5' can stand either for '|5|' or for '+5'), and the widely found double use of words for 'human being' to also denote a member of the tribe speaking the language (e.g. in the Maidu language, *majdy* means 'human being' or 'Maidu person', contrasting with the 'marked' expressions *kombom majdy* 'Yana person', *wolem majdy* 'white person', and so on).

The general property of 'human thinking' that is involved here is the fact that our communication is guided by highly general pragmatic principles of the sort that have been discussed in the neo-Gricean literature (e.g. Horn 1984, Levinson 2000). Specifically, in cases like *majdy* the highly general Principle of Informativeness ('I-Principle') is relevant, which allows the hearer to enrich the interpretation of an utterance if no further specification is given by the speaker. Thus, *if p, q* can be interpreted as communicating by implicature 'if and only if p, q', and *p and q* can be interpreted as intending 'p and therefore q'. Similarly, in the context of Maidu society, the word *majdy* 'human being' will ordinarily be understood as referring to a Maidu person, because if the speaker had unexpectedly intended a non-Maidu, she would of course have used a special expression. The Maidu word *majdy* need not be polysemous or bi-functional at all. It probably just means 'human being', and the Maidu language simply lacks a conventional word for 'Maidu person'. There is thus no need to appeal to markedness in such cases. We do need Gricean pragmatic principles of utterance interpretation to understand how language functions and is structured, but these principles are far more general.

Something very much like the distinction between semantic content (what is said) and pragmatic implicatures (what it communicates) is already present in Jakobson's (1932[1971: 3f.]) discussion of the conditions under which the zero-meaning of the unmarked member (representing the entire category, e.g. Russian *osël* for 'donkey in general') occurs as opposed to the minus-meaning (representing *par excellence* the opposite member to the marked member, e.g. *osël* for 'male donkey').¹⁸ He says that the minus-meaning is just a situationally conditioned use of the category (i.e. an implicature) which should not be confused with its general meaning (i.e. its semantic content). However, he does not specify under what conditions the minus-meaning arises. Waugh (1982: 304) says that it occurs 'in individuating contexts and with an implicit or explicit contrast between marked and unmarked'.

[18] The terms 'zero-meaning' and 'minus-meaning' are from Waugh (1982).

Jakobson gives the following example, where the interpretation ‘male donkey’ (the minus-meaning) seems to arise because of the contrastive context.

- (3) Èto oslíca? Nét, osěl.
 ‘Is that a female donkey? No, a male donkey.’

However, in most lexical pairs in which one member is semantically more specific, no such minus-meaning arises in contrastive contexts:

- (4) (a) Is that a bus? *No, a vehicle.
 (b) Is that a microorganism? *No, an organism.
 (c) Is that horse galloping? *No, it’s moving.

In the neo-Gricean tradition, it is generally assumed that no special contexts are needed to trigger Informativeness-based implicatures like “human being” (in a Maidu context) implicates “Maidu person”. On the contrary, these implicatures are canceled only under specific conditions, namely when they are explicitly contradicted (‘That’s a human being, but not from our tribe’). Thus, informativeness-based implicatures cannot be the whole story for pairs like Russian *oslíca/osěl* (or the analogous English *lioness/lion*).

Now as is well known, pragmatic inferences (implicatures) may become so common that at some point they are reanalyzed as being not just part of what is communicated, but part of what is said, i.e. they become non-cancelable. Consider, as an example, the word *America*. Originally this referred to the entire continent in the western hemisphere that Europeans had become aware of after 1492. But English speakers of course used *America* primarily for the part of the continent that was settled from England, and nowadays it has become restricted to the United States of America. It is no longer possible to cancel this enriched meaning in English (**I’m from America, more specifically from South America*). The enriched meaning has become conventionalized.

Similarly, Witkowski & Brown (1983: 571f.) report on ‘markedness reversal’ in the word for ‘deer’ (*čih*) in several Mayan languages of Chiapas (Mexico). When sheep were introduced by the European invaders, these were referred to as ‘cotton deer’ (Tzeltal: *tunim čih*, Tzotzil: *tušnok’ čih*). As sheep became more important in Chiapas culture, there was less and less need for the additional specification, and the simple word *čih* is now used for ‘sheep’. To refer to ‘deer’, Tzotzil now uses the optional modifier *teʔtikil* ‘wild’ (*teʔtikil čih*), so in this language *čih* can still be said to be vague with respect to the sheep vs. deer distinction. In Tenejapa Tzeltal, by contrast, the modifier is now obligatory, and *čih* alone can only mean ‘sheep’. What was originally an Informativeness-based implicature is now the conventional meaning.

In order to explain why it is possible to say *Èto oslíca? Nét, osěl* in Russian and *Is that a lioness? No, a lion* in English (for many speakers), we have to assume some conventionalization. The pair *lioness/lion* has similarities both to the pair *female frog/frog* (where we have a simple hyponym/hyperonym

relation) and to the pair *princess/prince* (where the two members are clearly mutually exclusive). The suffix *-ess* may have functioned just like the adjective *female* at an earlier time, so that the relation was one of pure hyponymy. But now the suffix *-ess* is obligatory when referring to females, and *prince* has become restricted to males. This is much like the restriction of *America* to the United States, and the obligatoriness of the modifiers *North* and *South* for referring to the purely geographical concepts. The pair *lioness/lion* seems to show incipient conventionalization.

Now one might say that it is precisely for term pairs with incipient conventionalization like *osël/oslica*, *lion/lioness* that the special term 'markedness' is appropriate. But when we consider additional behavioral criteria besides contrastive predication, we see that different kinds of term pairs show quite different behavior, and we need more than the three categories 'no conventionalization', 'incipient conventionalization', and 'full conventionalization'. Let us call the two members of a term pair 'term-1' and 'term-2'.

The most permissive criterion seems to be that of heterogenous plurals, i.e. the plural of term-1 refers to a group that includes members described by term-2. Thus, *lions* can refer to a group of both lionesses and male lions, and in German, *Ärzte* (the plural of *Arzt*) can refer to a group of male physicians (*Arzt*) and female physicians (*Ärztin*). In Spanish, even *padres* (plural of *padre* 'father') can refer to a pair consisting of a father and a mother (i.e. the parents).

However, another criterion is the use of term-1 for term-2 in specific reference. This is possible for *lion/lioness*, but not for German *Arzt/Ärztin* or Spanish *padre/madre*:

- (5) (a) I saw a lion. (It was a lioness.)
- (b) *Ich sah einen Arzt. (Es war eine Ärztin.)
 'I saw a doctor. It was a female doctor.'
- (c) *Vi a un padre. Fue una madre.
 'I saw a father/parent. It was a mother.'

To test whether we are dealing with true hyponymy, we can use a frame such as 'A term-2 is a (particular kind of) term-1'. By this criterion, we may still have hyponymy with *Arzt/Ärztin*, but *cow/bull* is not hyponymy anymore (let alone *king/queen*):

- (6) (a) A lioness is a female lion.
- (b) ?Eine Ärztin ist ein weiblicher Arzt.
- (c) *A bull is a male cow. (Lyons 1977: 309)
- (d) *A queen is a female king.

Yet other criteria, such as 'That 1 is a 2' (e.g. *That lion is a lioness/?That cow is a bull*), and the non-specific use of 1 for 2 (*I'm looking for a lion* may imply that I would also be satisfied with a female lion, cf. Doleschal 1995), may give still different results. And some lexical items may behave totally

TERM 1	TERM 2	CONTRASTIVE PREDICATION EXCLUDED (‘this is a 1, not a 2’)	SPECIFIC USE OF I FOR 2 ‘I saw a 1’ (meaning ‘I saw 2’)	HYPONIMY (‘2 is a (particular) kind of 1’)	NONSPECIFIC USE OF I FOR 2 (‘I’m looking for a 1’ includes 2)	HETERO- GENEOUS PLURALS (‘is’ can stand for ‘is and 2s’)
			‘that 1 is a 2’			
<i>vehicle</i>	<i>bus</i>	Y	Y	Y	Y	Y
<i>soldier</i>	<i>general</i>	??	Y	Y	Y	Y
<i>lion</i>	<i>lioness</i>	??	Y	Y	Y	Y
<i>cow</i>	<i>bull</i>	–	?	??	?	Y
<i>house</i>	<i>church</i>	–	?	??	??	?
<i>Arzt</i>	<i>Ärztin</i>	–	–	–	?	Y
<i>padre</i>	<i>madre</i>	–	–	–	–	Y
<i>man</i>	<i>woman</i>	–	–	–	–	(Y)
<i>king</i>	<i>queen</i>	–	–	–	–	–

Table 3

Different term pairs and different potential ‘markedness’ criteria

idiosyncratically: *man* is the only word in English that can have an inclusive sense only in a generic context (also the inclusive sense in the plural seems to be possible only in generic contexts). The behavior of these pairs and a few more can be summarized as in table 3.¹⁹ This table shows an interesting pattern, but saying that term-1 is ‘unmarked’ and term-2 is ‘marked’ does not help much to describe or understand the pattern. Language learners need to acquire much more specific properties of these words, so linguistic descriptions cannot content themselves with a simple bifurcation either. And to really explain what is going on, we need to refer to a variety of factors, among them clearly frequency of use: in the pair *dog/bitch*, *bitch* has a much lower proportional frequency than *queen* has in the pair *king/queen*, so it is not surprising that it behaves more like a hyponym of *dog*. (We find a different frequency effect in pairs that have nothing to do with gender: in pairs like *vehicle/bus*, term-2 is relatively more frequent, whereas in pairs like *soldier/general*, term-1 is relatively more frequent. Again, the frequency is apparently relevant to explaining the difference in behavior.)

In grammatical semantics, too, hyponymy or semantic specification is insufficient as an explication of semantic markedness. As in lexical semantics, or even more so, the neutralizing contexts are highly specific and must be

[19] There is a fair amount of between-speaker variation in English. I am grateful to Jim Blevins, Juliette Blevins, Bernard Comrie, Grev Corbett, Jeff Good, Orin Gensler, and David Kamholz for sharing their intuitions with me.

learned for each language and for each category separately. In English, it is not possible to use the present tense in many past contexts even when the past tense is redundant (**Yesterday I lose my keys*), and similarly the singular is not usable for plural reference even where there is no need for it (**three book*). In other languages, these sentences are perfectly normal (in fact, many languages require the singular in combination with numerals). Similarly, the (allegedly marked) perfective aspect is obligatory in many contexts in Russian, and the details of the use of the perfective/imperfective aspects are highly complex and language-particular.

We can conclude that grammars do not become significantly more elegant if we build markedness into them, and it has not been shown that any concept of markedness helps us understand whatever general patterns may lie behind the cross-linguistic and intra-linguistic diversity in lexical and grammatical semantic contrasts.

What clearly remains valid is the cross-linguistic tendency for distributional asymmetries of the kind discussed here to correlate with overt coding and frequency: the more frequent term of an opposition tends to be coded with less material or zero (see section 4), and the more frequent term tends to be more widely distributed. But the wider distribution of terms like *lion*, *cow*, *dog* is not in the first instance due to their greater frequency – rather, it is due to their wider meaning, which is itself ultimately due to the lower frequency of the opposite meaning.

6. REPLACING ‘PHONOLOGICAL MARKEDNESS’

Markedness in phonology has recently been critically discussed in some detail by Hume (2004) (‘Deconstructing markedness’) and by Blevins (2004) (see also Gurevich 2001). Blevins states that in her approach, which emphasizes the role of regularities of sound change for explaining synchronic grammars (‘Evolutionary Phonology’),

there is no clear role for markedness within synchronic phonology. Absolute universals and universal tendencies in sound patterns emerge from general pathways of language change, and have no independent status in the grammar ... there is a great deal of empirical evidence against the direct incorporation of markedness into synchronic grammars. (Blevins 2004: 20)

Hume and Blevins downplay markedness because it is too abstract a notion, and they would rather explain the corresponding phenomena directly in terms of phonetics, language use and language change.²⁰

[20] Cf. also formulations such as the following from the closing paragraph of Crosswhite (2004: 226): ‘there is no such thing as a monolithic concept of markedness ... In effect, the

But even those who are more interested in the most abstract properties of the cognitive code have good reasons to be suspicious of markedness explanations. Thus, Hale & Reiss (2000: 162), who take an antifunctionalist, strictly Chomskyan stance, observe:

[M]any of the so-called phonological universals (often discussed under the rubric of markedness) are in fact epiphenomena deriving from the intersection of extragrammatical factors like acoustic salience and the nature of language change.

Thus, in phonology, the relevance of ‘markedness’ has already been questioned quite radically by prominent phonologists, so that my treatment of phonological markedness can be relatively brief.

In the list of senses in section 2 I started with ‘specification for a phonological distinction’ (sense 1) because this was the original sense of the term ‘marked’. The only criterion admitted by Trubetzkoy for identifying the ‘mark-bearing’ term was language-particular neutralization (a kind of restricted distribution, sense 10). But already Jakobson (1941) associated asymmetries in phonological oppositions with differences in phonetic complexity and ‘contrastive optimality’ (see 1941[1962: 373f.]), so he was the first to introduce the notion of markedness as a universal property of sounds that is rooted in phonetic difficulty (sense 4). In the same work, he also noted that markedness is reflected in typological patterns (languages have marked phonemes only if they have the corresponding unmarked phonemes, sense 9), and in the order of acquisition and loss of phonological contrasts. Greenberg (1966) was not interested in the nature of phonological systems as such, but he stressed the apparently universal correlation between neutralization, typological implication, text frequency, and two other phenomena: allophonic variation and phonemic differentiation. Thus, the following observable phenomena should all identify the same properties as marked and unmarked:

- (i) **NEUTRALIZATION**: In neutralization, the unmarked term stands for the marked term (e.g. in German, the voiced-voiceless opposition is neutralized in obstruents syllable-finally in favor of voiceless obstruents).
- (ii) **TYPOLOGICAL IMPLICATION**: If a language has the marked term, it has the unmarked term.
- (iii) **FREQUENCY**: The unmarked term is more frequent than the marked term.
- (iv) **ALLOPHONIC VARIATION**: The unmarked term shows greater allophonic variation than the marked term (e.g. Nootka unglottalized stops have aspirated and unaspirated allophones, whereas glottalized stops have just a single allophone).

phonological concept of markedness has been replaced by phonetic considerations, which are encoded in phonology using phonetically motivated constraints’.

- (v) **PHONEMIC DIFFERENTIATION**: The unmarked value may be more differentiated, i.e. there may be more phonemes having this value (e.g. languages may have more nonnasal vowels than nasal vowels, and they never have fewer).

Greenberg thus saw phonological markedness as a multidimensional correlation, just like morphological and lexical markedness (see section 2.4, section 4.5 above), and he emphasized the similarities between the characteristics in (i)–(v) and those that he identified in morphology and the lexicon (see also Croft 1990: ch. 4). However, he also noted that frequency cannot be the single explanatory factor for the correlation, unlike in grammar and lexicon, and Croft (2003: 117–120) no longer includes the phonological correlations in his markedness concept.

In more recent work on phonological markedness (e.g. McCarthy & Prince 1994, de Lacy 2002, Rice 2003), two other markedness correlates have become prominent:

- (vi) **INSTABILITY IN ASSIMILATION**: Unmarked segments assimilate more easily than marked segments (e.g. Korean *kotpalo* → *koppalo* ‘straight’, but *papto* ‘rice also’ does not become **patto*; Rice 2003: 395).
- (vii) **EMERGENCE OF THE UNMARKED**: In epenthesis, only unmarked segments appear (e.g. [ə], [t]).

As in the case of morphology and semantics, there is some question about whether the correlations actually hold even as statistical tendencies,²¹ and most of the available evidence for (or against) them is anecdotal. But to the extent that they do hold, it is clear that it is more promising to try to account for the various correlating dimensions individually than to reify the correlations as ‘markedness’ (perhaps encoded as such in individual grammars or even in Universal Grammar, i.e. in the cognitive code for grammars).

As Croft (2003: 120) notes, the explanation for the correlations ‘is likely to be found in articulatory and acoustic phonetics’, so let us start with ‘markedness as phonetic difficulty’ (sense 4, section 2.2.1 above). Just as conceptual difficulty is causally related to rarity in texts (see section 4.4), phonetic difficulty or complexity is closely related to rarity. Phoneme systems are more efficient (in particular, more speaker-friendly) if they privilege articulatorily simple segments than if they privilege more complex segments.²² While the hearer would have no gain, the speaker would suffer if a language used its ejective or palatalized consonants more often than its plain consonants. This

[21] For instance, de Lacy (2002: 6) claims that ‘there are no asymmetries in assimilation and coalescence’, contrary to what appears to be the majority opinion.

[22] The same is true for other phonological units such as syllables, but to keep the discussion simple, I focus on segments here.

is thus an economy consideration, similar to the explanation of the correlation between shortness and frequency (see section 4.3). It seems that the causal direction can go both ways (as with conceptual difficulty, section 4.4): articulatory complexity can cause rarity (because speakers will unconsciously prefer expressions with simple segments for items that occur frequently in discourse, and this will find its reflection in the system), and frequency can cause articulatory simplicity (because when contrasts are given up after becoming redundant, this often involves articulatory simplification). The same is true for perceptual difficulty: perceptually salient (=easy-to-perceive) segments will be preferred by speakers in discourse and thus become frequent in the system. According to Hume (2004), ‘sound sequences with richer cues [i.e. perceptually more salient sound sequences] tend to be more stable in a language and consequently, typically occur in more words than those with weaker cues’.

The greater distributional range of some segments compared to others (including neutralization) is also mostly due to phonetic difficulty. German voiced obstruents do not occur syllable-finally because, for phonetic reasons, voice contrasts are harder to maintain in this position (see, e.g., Blevins 2004: 103–106). In Bulgarian unstressed syllables, the contrast between /i/–/e/, and between /u/–/o/, is neutralized in favor of the high vowel (Trubetzkoy 1939: 73), because /i/ and /u/ are more salient and contrast better with /a/ than /e/ and /o/ (Crosswhite 2004). The distributional restrictions on phonetically difficult segments contribute further to the higher frequency of the easier segments, but they hardly suffice to explain the differences (cf. the analogous discussion for meaningful elements in section 4.2) because more complex segments are generally rarer also in non-neutralized positions (e.g. palatalized consonants in Russian are rarer also before vowels, voiced obstruents are rarer in German also in syllable-initial position).

Likewise, typological implication can be explained by phonetic difficulty if one assumes the generally accepted regularity that the ability to perform a more difficult task implies the ability to perform easier tasks. Speakers who master a voiced uvular stop [g] should also be able to pronounce the velar [g]. And if for some reason /g/ disappeared from the language, /G/ would probably soon be pronounced [g] because there would be no need for the extra articulatory effort required to pronounce [G]. Similarly, if hearers can perceive a relatively non-salient contrast like that between [ø] and [e], they will also be able to distinguish between [i] and [e] (thus leading to the typological law that an /ø/ in the system implies an /e/); and if for some reason the /e/ disappeared, /ø/ would presumably shift to /e/, thus making the contrast maximally salient.²³

[23] A somewhat different, diachronic explanation for such implications is proposed by Greenberg (1966) and Bybee (2001: 202).

The other correlates of phonological ‘markedness’ can be explained by frequency of use and predictability (Hume 2004). Rare segments are unexpected, i.e. hard to predict, so they resist assimilation. By contrast, ‘the greater the predictability of an element, the less information content it has and ... the more expendable it is’ (Hume 2004: section 5.1). The explanation for instability in assimilation is thus very similar to the explanation given for articulatory simplicity above. Hume also argues that the appearance of ‘unmarked’ vowels in epenthesis is due to their higher frequency: it is ‘a segment’s predictability in a given context that is crucial in determining whether it will be perceived as the epenthetic vowel’ (Hume 2004: section 5.3).

Since frequency may differ to some extent across languages, for a variety of reasons, the frequency effects are sometimes different in different languages. Thus, while English-speaking children usually acquire /k/ later than /t/, Japanese-speaking children have more difficulty with /t/ than with /k/. The reason is simply that Japanese has a higher frequency of /k/ (Hume 2004: section 5.2, citing work by Mary Beckman and colleagues).

Thus, phonological contrasts are similar to semantic contrasts in that we appear to find evidence for cross-linguistic correlations between a number of logically independent dimensions. But as in the case of meaningful (lexical and morphological) categories, ‘markedness’ is not needed, because the correlations can be explained directly with reference to phonetic factors and frequency of use. An abstract markedness notion, whether intended as a claim about the cognitive code, as a convenient metagrammatical term, or as an explanatory concept in an autonomous theory of markedness (see section 3), only makes it harder to understand the cross-linguistic and language-particular patterns.

7. REPLACING ‘MARKEDNESS AS MORPHOLOGICAL DIFFICULTY’

Natural morphologists have made a number of claims about the ‘naturalness’ (=unmarkedness) of morphological structures. In particular, they have claimed that the better a structure fares with respect to a number of ‘naturalness parameters’, the more frequent it will be within and across languages, and the more robust it will be in psycholinguistic evidence and in language change. The main universal naturalness parameters are constructional iconicity, uniformity, and transparency (see section 2.2.2). I will discuss each of these in turn, and then go on to discuss ‘system-dependent naturalness’.

Constructional iconicity essentially means that ‘what is “more” semantically should also be “more” constructionally’ (Mayerthaler 1981: 25, cf. Wurzel 1994: 2592). This was already discussed toward the end of section 2.4, where we saw that the basic generalization is easily explained by economy: what is used more frequently is shorter in any rational communication system. No appeal to an iconicity principle is needed (see also Haspelmath, in preparation). The natural morphologists have proposed a more fine-grained

scale between maximal iconicity (when segments are added, as in English *boy/boys*), minimal iconicity (when the two categories just differ in the shape of the stem, as in *goose/geese*), non-iconicity (as in *sheep/sheep*), and counter-iconicity (as in Welsh *plu-en* 'feather'/*plu* 'feathers'). That counter-iconicity should be dispreferred is also predicted by frequency-based economy, and the obvious problem with 'non-iconicity' is its uninformativeness (though this can arise only in languages that usually do have a clearly coded distinction; quite a few languages have no nominal singular-plural distinction at all). More interestingly, the apparent preference for 'maximal iconicity' over 'minimal iconicity' may well be an artifact of historical pathways. Bybee & Newman (1995) argue that stem changes (as in *goose/geese*) are just as easy to learn and process as affixes, and that the fact that they are generally rarer both within and across languages has a diachronic explanation: morphology generally arises via the grammaticalization of earlier lexical items, and hence stem changes do not have the same opportunities as affixes to arise in the first place. Since they are generally rare within a language, it is expected that they should often be under pressure to be leveled on the analogy of the more widespread affixational patterns, and that they do not often become fully productive. The low type frequency of stem changes can also be made responsible for the difficulties they present in acquisition and processing. However, once a language has a fully productive system of stem changes, this may well be stable for millennia, as the Semitic languages show.

Uniformity means that stem morphemes show no allomorphy. The widely observed tendency to eliminate stem alternations (as in the change from *brethren* to *brothers*, or from *hou[z]es* to *hou[s]es*) is attributed to this markedness/naturalness principle. Mayerthaler (1987: 49) claims that it is 'based upon biologically given preferences of the human brain', and links it to the perceptual preference for 'object constancy'. But it is possible to be more specific. As Bybee (1985: 119-123) noted, leveling of stem alternations is highly frequency-sensitive: 'the proposal that infrequently-used forms fade [from memory] accounts for the tendency to regularize infrequent irregular forms, for an irregular form that is not sufficiently reinforced will be replaced by a regular formation' (1985: 119). Thus, stem alternations in highly frequent words (like *does*, *says*) can be very stable, and they do not make the morphology any more difficult for language users (cf. Werner 1989).

Transparency is the tendency for morphological paradigms to avoid homonymy. At some level such a principle may well exist, but again its effects are not independent of frequency: as we saw in section 4.5, the rarer morphological categories tend to exhibit less differentiation, i.e. more syncretism/homonymy, than the more frequent categories, the reason being that the distinctions are harder to remember in the rare categories. Thus, although transparent paradigms make life easier for the hearer, the learner may well find them harder or 'unnatural'. A single dimension of naturalness/unmarkedness is clearly insufficient here.

Finally, Wurzel (1984, 1987) also extends the notion of unnaturalness/ markedness to language-particular phenomena: ‘what is “better” or “worse” in the morphology for speakers, what is more marked or less marked, depends crucially on what is normal in their system’ (Wurzel 1994: 2594). This explains why in some cases more homonymy is introduced. For example, in German the dative form *dem Bären* ‘to the bear’ is giving way to *dem Bär-Ø* in the colloquial language, although it runs counter to constructional iconicity. But the predominant nominal pattern is the one without a suffix in the dative case (e.g. *dem Wolf* ‘to the wolf’), and so the minority paradigm of *Bär/Bären* simply joins the majority paradigm of *Wolf*. As Wurzel recognizes, the crucial factor here is type frequency (although he calls it ‘normalcy’, as if it were not easily measurable). As Bybee (1995) (among many others) emphasizes, type frequency is one of the key factors determining morphological productivity, for well-understood psychological reasons. Wurzel (2000) (and similarly in earlier work, cf. 1984: 136ff.) also claims that words belonging to minority classes are ‘marked’ in that they require an additional lexical specification.²⁴ Thus, *Bär* (in the standard language) requires the lexical specification ‘*n*-declension’, whereas *Wolf* requires no such specification. Thus, ‘morphological difficulty’ here reduces to remembering an additional piece of information about a word. But in a more psychologically oriented model such as Bybee’s, there is no simple contrast between specification and non-specification. Instead, the organization in memory of both words and patterns is highly dependent on their use, especially their frequency, and in some cases there is every reason to believe that even predictable information is stored (highly frequent regular words must have separate storage of their inflected forms, otherwise it would be impossible to explain how they could become irregular, as in the case of *have/has*).

Thus, the tendencies that natural morphologists have tried to explain by invoking a notion of markedness/unnaturalness as difficulty can all be explained by independently motivated factors. ‘Markedness’ is superfluous.

8. REPLACING THE SIX ROLES OF MARKEDNESS

Let me now address the various roles that markedness has been given by linguists (cf. section 3), and briefly say why markedness is not necessary at each of these levels. Actually, I will only mention the first five roles of section 3, because the sixth role has already been dismissed.

- (i) **MARKEDNESS AS REPRESENTED IN PARTICULAR GRAMMARS.** This concept has not been prominent since Trubetzkoy (even Jakobson later came to emphasize the universal nature of markedness asymmetries). Language

[24] In this way, Wurzel’s use of ‘markedness’ in the context of inflectional classes comes to resemble senses 1 and 2 above.

descriptions (such as reference grammars or dictionaries) get by easily without reference to markedness. Even in purely theoretical descriptions, it turns out on closer examination that markedness does not lead to greater elegance of description. For example, Lyons (1977: 308) argues that the twofold function of words like *dog* (in pairs like *dog/bitch*) ‘should not be treated as an instance of polysemy’, because such words are widespread throughout the English lexicon, and one would expect polysemy to idiosyncratically target particular words. However, as we saw in section 5, word pairs of this kind can in fact behave in a wide variety of ways, so a single markedness notion does not really help.

- (ii) **MARKEDNESS AS PART OF THE COGNITIVE CODE** (‘UNIVERSAL GRAMMAR’). Generative grammarians have typically tried to do two things simultaneously: characterize the innate cognitive prerequisites for acquiring a language (i.e. the cognitive code), and account for the limits on cross-linguistic variation. Thus, Chomsky & Halle (1968: ch. 9) tried to refine their formal descriptive framework in such a way that the difference between a cross-linguistically widespread (‘natural’) pattern (e.g. a vowel inventory /i e a o u/) and an unusual pattern (e.g. a vowel inventory /y œ a ʌ i/, Chomsky & Halle 1968: 402) would fall out from it. With their ‘marking conventions’, they effectively built the explanation for the cross-linguistic patterns into their descriptive framework. They did not even consider the possibility that the cross-linguistic patterns could have totally different, extralinguistic explanations, unrelated to the limitations of the cognitive code (just as organisms are constrained much more by factors other than the limitations of the genetic code; see Haspelmath 2004a). This was already pointed out by Anderson (1974: 293), who noted that phonological structures ‘closely mirror the mechanisms of speech physiology insofar as we understand them’. He continues:

One must start from the description and search for an explanation, rather than attempting to make the explanation shape the description. A correct substantive theory of [naturalness issues], then, is not to be sought in a restructuring of phonological description as is implicit in [Chomsky & Halle’s] marking conventions.

About three decades later, the currently most popular generative framework, Optimality Theory (OT), still tries to apply Chomsky & Halle’s (1968) old strategy of building explanations into descriptions. McCarthy (2002: 2) writes:

One of the most compelling features of OT, in my view, is the way that it unites description of individual languages with explanation in language typology ... the grammar of one language inevitably incorporates claims about the grammars of all languages.

But for the same reasons, this strategy of OT cannot work: ‘markedness’ phenomena are ultimately due to substantive factors, and abstract innate constructs such as OT constraints are simply not the right place for explanation (see especially Blevins (2004: ch. 9) for a critique of OT in phonology, and Haspelmath (1999b) for a more general critique).

- (iii) **MARKEDNESS AS A METAGRAMMATICAL CONCEPT** (as used by Greenberg (1966) and Croft (1990, 2003) for a multidimensional correlation). Here I would argue that retaining a separate label, even when it is recognized that frequency differences explain the correlation, only leads to confusion.²⁵ The same is true for ‘markedness’ in the sense of ‘narrowly distributed’ or ‘non-default’. Metagrammatical terms should be maximally transparent, and there should be no need for linguistics-specific terminology at this level of generality.
- (iv) **MARKEDNESS AS AN EXPLANATORY CONCEPT.** As Wurzel (1998) and other natural morphologists recognize, the ultimate explanation for regularities of language structure is in terms of substantive factors outside the language system. If one posits an abstract intermediate ‘explanatory’ level between the phenomena and the real explanatory factors, one bears the burden of proof that such a level is needed. I have argued (section 7) that it is superfluous.
- (v) **MARKEDNESS AS A GENERAL PROPERTY OF HUMAN CULTURE.** Frequency asymmetries and expectations based on such asymmetries also occur outside of language. Mayerthaler (1981: 48) cites the example of clothing habits on beaches: normally people wear bathing suits, so naked bathing is the marked case. On nudist beaches, we find markedness reversal – wearing a bathing suit becomes marked, i.e. unexpected, and therefore remarkable. Jakobson (in his 1930 letter to Trubetzkoy) mentions the Soviet policy shift from the default assumption that people are pro-Soviet to the default expectation that people are against the regime (‘all those who are not with us are against us’). This example shows that default expectations can be influenced by factors other than frequency asymmetries (e.g. paranoid imagination of frequency asymmetries); but such factors seem irrelevant for language, and the concept of ‘default expectation’ seems fully satisfactory to describe such situations. More interesting is the claim in Moravcsik & Wirth (1986) that non-linguistic cultural items such as food and clothing tend to show the same correlation between frequency,

[25] This is implicitly recognized by Croft (2003: 87), who replaces the chapter heading ‘Markedness’ of Croft (1990) by ‘Typological markedness’. However, this latter term does not improve the situation greatly, because it has been used widely in the sense of typological implication (sense 9, section 2.3.3), and this is not even one of the correlates that Croft (2003) discusses.

diversity and simplicity that is seen in language: everyday clothing is more frequent, comes in more different kinds, and is simpler than festive attire (and similarly for food). Here the simplicity cannot be due to frequency, but in fact there is no correlation: zero clothing (or food), which would be predicted to be even more frequent than simple clothing (or food), is in fact even rarer than festive attire (or holiday food). And the correlation with diversity does not hold either: many people have a much greater range of holiday clothes, contrasting with uniform working clothes. For food, if there is a correlation at all, it is explainable by memory limitations (cf. the explanation for reduced inflectional differentiation in rare categories, section 4.5(iii)): recipes that are used rarely are less easily remembered. Those who cook from cookbooks (or go to a restaurant) on special days do not have to suffer from this unfortunate restriction on food diversity.

9. CONCLUSION

Some writers on markedness complain that ‘capturing exactly what markedness means is by no means a straightforward task’ (Rice 2003: 390). In this paper, I have not set out to find ‘the’ meaning of ‘markedness’, but I have attempted to identify and characterize a range of different senses (and contexts) in which the term is used in linguistics. While Rice (2003: 419) believes that ‘[m]arkedness is something about which linguists come to have strong intuitions’, I find little evidence for the hope that these intuitions overlap sufficiently to allow us to agree on a common textbook standard anytime soon. On the contrary, it seems that the ‘intuitive’ shared sense of ‘marked/unmarked’ is not distinguishable from the sense of everyday words like *uncommon/common*, *abnormal/normal*, *unusual/usual*, *unexpected/expected*. Apart from the larger class of markedness as abnormality (section 2.3), we also find markedness as complexity (section 2.1) and as difficulty (section 2.2), but since complexity and difficulty typically lead to lower frequency, abnormality is in effect what all markedness senses share. But we do not need a technical linguistic term for abnormality/uncommonness/unusualness/unexpectedness. Simple everyday concepts should be expressed by simple everyday words.

In sections 4–7, I have addressed the various senses of markedness individually, trying to show for each sense how we can deal with the observed phenomena in an insightful way without invoking a markedness notion. A summary of the conclusions is provided in table 4.

I conclude that linguists can dispense with the term ‘markedness’ and many of the concepts that it has been used to express. It can be readily replaced by other concepts and terms that are less ambiguous, more transparent and provide better explanations for the observed phenomena. At the very least, non-use of this term by authors should become readers’ default

TYPE OF MARKEDNESS	DOMAIN	SALIENT REPRESENTATIVE	WHAT TO REPLACE IT WITH	DISCUSSED HERE IN ...
1. Markedness as specification for a phonological distinction	phonemes	Trubetzkoy 1939	detailed phonetic and distributional description	section 2.1.1
2. Markedness as specification for a semantic distinction	lexical items, grammatical categories	Jakobson 1932	detailed semantic description and pragmatic analysis	section 5
3. Markedness as overt coding	morphosyntactic categories	(passim)	‘overt coding/zero coding’	section 4.3
4. Markedness as phonetic difficulty	phonological/phonetic categories	Hayes & Steriade 2004	detailed study of phonetic factors	section 6
5. Markedness as morphological difficulty/unnaturalness	morphological patterns	Dressler et al. 1987, Wurzel 1998	general principles of mental organization of words (especially frequency differences)	section 7
6. Markedness as conceptual difficulty	(grammatical) conceptual categories	Givón 1991, 1995	conceptual difficulty (sometimes due to rarity of occurrence)	section 4.4
7. Markedness as rarity in texts	any linguistic element or pattern	Greenberg 1966	‘rarity in texts’	section 4.2

8. Markedness as rarity in the world	anything that people can perceive	(passim)	‘rarity in the world’	section 4.2
9. Markedness as typological implication or cross-linguistic rarity	phonological categories, syntactic patterns	Jakobson 1941, 1963	‘typological implication, cross-linguistic rarity’	section 6
10. Markedness as restricted distribution	phonological and grammatical categories, lexical items, syntactic patterns	(passim)	1. restricted distribution vs. unrestricted distribution 2. specifically defined distribution vs. default distribution	section 2.3.4
11. Markedness as deviation from default parameter setting	parametric options	Chomsky 1981	‘deviation from default parameter setting’	section 2.3.5
12. Markedness as a multi-dimensional correlation	grammatical categories, lexical items (and perhaps phonological categories)	Greenberg 1966, Croft 1990	rarity/frequency in texts and its consequences	section 4.5

Table 4

The twelve markedness senses, their domains and salient representatives, and what to replace them with

expectation, so that authors who feel that they need the term are obliged to motivate its use.

REFERENCES

- Aissen, Judith (1999). Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory* 17, 673–711.
- Aissen, Judith (2003). Differential object marking: iconicity vs. economy. *Natural Language & Linguistic Theory* 21, 435–483.
- Andersen, Henning (1989). Markedness theory – the first 150 years. In Mišeska Tomić (ed.), 11–46.
- Andersen, Henning (2001). Markedness and the theory of linguistic change. In Andersen, Henning (ed.), *Actualization: linguistic change in progress*. Amsterdam: John Benjamins. 21–57.
- Anderson, Stephen R. (1974). *The organization of phonology*. New York: Academic Press.
- Anderson, Stephen R. & Lightfoot, David W. (2002). *The language organ: linguistics as cognitive physiology*. Cambridge: Cambridge University Press.
- Archangeli, Diana B. (1988). Underspecification in phonology. *Phonology* 5.2, 183–207.
- Archangeli, Diana B. (1992). Markedness in phonology. In Bright (ed.), 2: 391–393.
- Archangeli, Diana B. (1997). Optimality Theory: an introduction to linguistics in the 1990s. In Archangeli, Diana B. & Pulleyblank, Douglas G. (eds.), *Optimality Theory: an overview*. Malden, MA: Blackwell. 1–32.
- Archangeli, Diana B. & Pulleyblank, Douglas G. (1994). *Grounded phonology*. Cambridge, MA: MIT Press.
- Ariel, Mira (2004). The rise (and potential fall) of reflexive pronouns. Ms., Tel Aviv University.
- Baayen, Harald, Burani, Cristina & Schreuder, Robert (1997). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of Morphology 1996*. Dordrecht: Kluwer. 13–33.
- Baker, Mark C. (1996). *The polysynthesis parameter*. Oxford: Oxford University Press.
- Battistella, Edwin L. (1990). *Markedness: the evaluative superstructure of language*. New York: Oxford University Press.
- Battistella, Edwin L. (1996). *The logic of markedness*. New York: Oxford University Press.
- Bechert, Johannes (1977[1998]). *Grammatische Kategorien: Affinität, Markiertheit und pragmatische Begründung*. Linguistic Agency University of Trier, Series A, Paper No. 47. [Reprinted in Bechert, Johannes (1998), *Sprache in Raum und Zeit*, vol. 1: *Kleine Schriften*. Tübingen: Narr. 151–63.]
- Bermúdez-Otero, Ricardo & Börjars, Kersti (2006). Markedness in phonology and syntax: the problem of grounding. *Lingua*.
- Bickerton, Derek (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7, 173–212.
- Blevins, Juliette (1995). The syllable in phonological theory. In Goldsmith (ed.), 206–244.
- Blevins, Juliette (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.
- Botha, Rudolf P. (1989). *Challenging Chomsky: the generative garden game*. Oxford: Blackwell.
- Brekhus, Wayne (1998). A sociology of the unmarked: redirecting our focus. *Sociological Theory* 16.1, 34–51.
- Bright, William (ed.) (1992). *International encyclopedia of linguistics*. New York: Oxford University Press.
- Brøndal, Viggo (1940). Compensation et variation, deux principes de linguistique générale. *Scientia* 9/10, 101–109. [Reprinted in Brøndal, Viggo (1943), *Essais de linguistique générale. Publiés avec une bibliographie des Œuvres de l'auteur*. Copenhagen: E. Munksgaard. 105–116.]
- Bybee, Joan L. (1985). *Morphology: a study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10.5, 425–455.
- Bybee, Joan L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan L. & Newman, Jean E. (1995). Are stem changes as natural as affixes? *Linguistics* 33, 633–654.

- Carnie, Andrew (2002). *Syntax: a generative introduction*. Malden, MA: Blackwell.
- Chomsky, Noam (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam (1982). *Noam Chomsky on the generative enterprise: a discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht: Foris.
- Chomsky, Noam & Halle, Morris (1968). *The sound pattern of English*. New York: Harper & Row.
- Comrie, Bernard (1986). Markedness, grammar, people and the world. In Eckman et al. (eds.), 85–196.
- Corbett, Greville G. (1983). Resolution rules: agreement in person, number, and gender. In Gazdar, Gerald, Klein, Ewan & Pullum, Geoffrey K. (eds.), *Order, concord and constituency*. Dordrecht: Foris. 175–206.
- Corbett, Greville G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. (2000). *Number*. Cambridge: Cambridge University Press.
- Corston-Oliver, Simon H. (2003). Core arguments and the inversion of the nominal hierarchy in Roviana. In Du Bois, John W., Kumpf, Lorraine E. & Ashby, William J. (eds.), *Preferred argument structure: grammar as architecture for function* (Studies in Discourse and Grammar 14). Amsterdam: John Benjamins. 273–300.
- Croft, William (1990). *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William (1996). Markedness and universals: from the Prague school to typology. In Jankowsky, Kurt R. (ed.), *Multiple perspectives on the historical dimensions of language*. Münster: Nodus. 15–21.
- Croft, William (2003). *Typology and universals* (2nd edn.). Cambridge: Cambridge University Press.
- Crosswhite, Catherine M. (2004). Vowel reduction. In Hayes et al. (eds.), 191–231.
- de Lacy, Paul (2002). *The formal expression of markedness*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (2004). *The Jarawara language of southern Amazonia*. Oxford: Oxford University Press.
- Doleschal, Ursula (1995). Referring to women. In Geiger, Richard A. (ed.), *Reference in multi-disciplinary perspective: philosophical object, cognitive subject, intersubjective process*. Hildesheim: Georg Olms Verlag. 277–298.
- Dressler, Wolfgang U., Mayerthaler, Willi, Panagl, Oswald & Wurzel, Wolfgang U. (1987). *Leitmotifs in Natural Morphology* (Studies in Language Companion Series 10). Amsterdam: John Benjamins.
- Dryer, Matthew S. (1995). Frequency and pragmatically unmarked word order. In Downing, Pamela & Noonan, Michael (eds.), *Word order in discourse*. Amsterdam: John Benjamins. 105–135.
- Dryer, Matthew S. (2004). Vacuous appeals to the number hierarchy. Paper presented at the Max Planck Institute for Evolutionary Anthropology, Leipzig, January 2004.
- Eckman, Fred R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning* 27.2. 315–330.
- Eckman, Fred R. (1996). A functional-typological approach to second language acquisition theory. In Ritchie, William C. & Bhatia, Tej C. (eds.), *Handbook of second language acquisition*. San Diego: Academic Press. 195–211.
- Eckman, Fred R. (2004). Universals, innateness and explanation in second language acquisition. *Studies in Language* 28.3. 681–703.
- Eckman, Fred R., Moravcsik, Edith A. & Wirth, Jessica (eds.) (1986). *Markedness*. New York: Plenum Press.
- Faingold, Eduardo D. (2003). *The development of grammar in Spanish and the Romance languages*. Basingstoke: Palgrave Macmillan.
- Fenk-Oczlon, Gertraud (1991). Frequenz und Kognition – Frequenz und Markiertheit. *Folia Linguistica* 25.3–4. 361–394.
- Frajzyngier, Zygmunt (2001). *A grammar of Lele*. Stanford, CA: CSLI Publications.
- Gair, James W. (1988). Kinds of markedness. In Flynn, Suzanne & O’Neil, Wayne (eds.), *Linguistic theory in second language acquisition*. Dordrecht: Kluwer. 225–250.
- Givón, T. (1991). Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language* 15.2. 335–370.

- Givón, T. (1995). Markedness as meta-iconicity: distributional and cognitive correlates of syntactic structure. In Givón, T. (ed.), *Functionalism and grammar*. Amsterdam: John Benjamins. 25–69.
- Goldsmith, John (ed.) (1995). *Handbook of phonological theory*. Oxford: Blackwell.
- Greenberg, Joseph (ed.) (1963). *Universals of language*. Cambridge, MA: MIT Press.
- Greenberg, Joseph (1966). *Language universals, with special reference to feature hierarchies* (Janua Linguarum, Series Minor 59). The Hague: Mouton.
- Grimshaw, Jane (2001). Optimal clitic positions and the lexicon in Romance clitic systems. In Legendre et al. (eds.), 205–240.
- Gurevich, Naomi (2001). A critique of markedness-based theories in phonology. *Studies in the Linguistic Sciences* 31.2, 89–114.
- Gvozdanović, Jadranka (1989). Defining markedness. In Mišeska Tomić (ed.), 47–66.
- Haegman, Liliane (1994). *Introduction to government and binding theory*. Oxford: Blackwell.
- Haider, Hubert (1993). Markiertheit in der generativen Grammatik. In Jacobs, Joachim, von Stechow, Arnim, Sternefeld, Wolfgang & Vennemann, Theo (eds.), *Syntax: an international handbook of contemporary research* (vol. 1). Berlin: Mouton de Gruyter. 635–645.
- Haiman, John (1983). Iconic and economic motivation. *Language* 59, 781–819.
- Haiman, John (1985). *Natural syntax*. Cambridge: Cambridge University Press.
- Hale, Mark & Reiss, Charles (2000). Substance abuse and dysfunctionality: current trends in phonology. *Linguistic Inquiry* 31, 157–169.
- Haspelmath, Martin (1999a). Explaining article-possessor complementarity: economic motivation in noun phrase syntax. *Language* 75.2, 227–243.
- Haspelmath, Martin (1999b). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18.2, 180–205.
- Haspelmath, Martin (2004a). Does linguistic explanation presuppose linguistic description? *Studies in Language* 28.3, 554–579.
- Haspelmath, Martin (2004b). Explaining the Ditransitive Person-Role Constraint: a usage-based account. *Constructions* 2004/2, 49 pp. [Free online journal, www.constructions-online.de]
- Haspelmath, Martin (in preparation). Frequency vs. iconicity in explaining grammatical asymmetries.
- Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hayes, Bruce, Kirchner, Robert & Steriade, Donca (eds.) (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Hayes, Bruce & Steriade, Donca (2004). Introduction: the phonetic bases of phonological Markedness. In Hayes et al. (eds.), 1–33.
- Horn, Laurence R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Schiffrin, Deborah (ed.), *Meaning, form, and use in context: linguistic applications*. Washington, DC: Georgetown University Press. 11–42.
- Horn, Wilhelm (1921). *Sprachkörper und Sprachfunktion*. Berlin: Mayer & Müller.
- Hume, Elizabeth (2004). Deconstructing markedness. *Berkeley Linguistics Society* 30.
- Hurch, Bernhard & Rhodes, Richard A. (1996). *Natural Phonology: the state of the art*. Berlin: Mouton de Gruyter.
- Ito, Junko & Mester, Armin (2003). *Japanese morphophonemics: markedness and word structure*. Cambridge, MA: MIT Press.
- Jackendoff, Ray (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jakobson, Roman (1932[1971]). Zur Struktur des russischen Verbums. In *Charisteria Guglielmo Mathesio*. Prague. 74–84. [Reprinted in Jakobson (1971), 3–15.]
- Jakobson, Roman (1939[1971]). Signe zero. In *Mélanges de linguistique offerts à Charles Bally*. Geneva. 143–152. [Reprinted in Jakobson (1971), 211–219.]
- Jakobson, Roman (1941[1962]). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. [Reprinted in Jakobson (1962), 328–401.]
- Jakobson, Roman (1957[1971]). Shifters, verbal categories, and the Russian verb. In Jakobson (1971), 130–147.
- Jakobson, Roman (1962). *Selected writings* (vol. I). The Hague: Mouton.
- Jakobson, Roman (1963). Implications of language universals for linguistics. In Greenberg (ed.), 263–278.

- Jakobson, Roman (1971). *Selected writings* (vol. II). The Hague: Mouton.
- Johns, Alana (1992). Deriving ergativity. *Linguistic Inquiry* 23, 57–87.
- Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.
- Kean, Mary-Louise (1992). Markedness. In Bright (ed.), 2: 390–391.
- Kilani-Schoch, Marianne (1988). *Introduction à la morphologie naturelle*. Berne: Peter Lang.
- Lakoff, Robin Tolmach (2000). *The language war*. Berkeley, CA: University of California Press.
- Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar*, vol. II: *Descriptive application*. Stanford, CA: Stanford University Press.
- Lazard, Gilbert (1989). Transitivity and markedness: the antipassive in accusative languages. In Mišeska Tomić (ed.), 309–331.
- Leech, Geoffrey, Rayson, Paul & Wilson, Andrew (2001). *Word frequencies in written and spoken English based on the British National Corpus*. Harlow: Longman.
- Legendre, Géraldine, Grimshaw, Jane & Vikner, Sten (eds.) (2001). *Optimality-theoretic syntax*. Cambridge, MA: MIT Press.
- Lehrer, Adrienne (1985). Markedness and antonymy. *Journal of Linguistics* 21, 397–429.
- Lenerz, Jürgen (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Narr.
- Levinson, Stephen C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Ludwig, Ralph (2001). Markiertheit. In Haspelmath, Martin, König, Ekkehard, Oesterreicher, Wulf & Raible, Wolfgang (eds.), *Language typology and language universals: an international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft) vol. I. Berlin: Mouton de Gruyter. 400–419.
- Lyons, John (1977). *Semantics* (2 vols.). Cambridge: Cambridge University Press.
- Mańczak, Witold (1970). Sur la théorie de catégories marquées et non-marquées de Greenberg. *Linguistics* 59, 29–36.
- Mayerthaler, Willi (1981). *Morphologische Natürlichkeit*. Wiesbaden: Athenaion. [English translation: Mayerthaler 1988.]
- Mayerthaler, Willi (1987). System-independent morphological naturalness. In Dressler et al. (eds.), 25–58.
- Mayerthaler, Willi (1988). *Naturalness in morphology*. Ann Arbor, MI: Karoma.
- McCarthy, John J. (2002). *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
- McCarthy, John J. & Prince, Alan (1994). The emergence of the unmarked: optimality in prosodic morphology. In Gonzalez, M. (ed.), *Proceedings of the North Eastern Linguistic Society* 24, 333–379.
- Mišeska Tomić, Olga (ed.) (1989). *Markedness in synchrony and diachrony*. Berlin: Mouton de Gruyter.
- Moravcsik, Edith A. & Wirth, Jessica (1986). Markedness – an overview. In Eckman et al. (eds.), 1–11.
- Müller, Gereon (1999). Optimality, markedness, and word order in German. *Linguistics* 37.5, 777–818.
- Newmeyer, Frederick J. (1992). Iconicity and generative grammar. *Language* 68, 756–796.
- Radford, Andrew (1988). *Transformational grammar: a first course*. Cambridge: Cambridge University Press.
- Radford, Andrew (2004). *Minimalist syntax: exploring the structure of English*. Cambridge: Cambridge University Press.
- Rice, Keren (2003). Featural markedness in phonology: variation. In Cheng, Lisa & Sybesma, Rint (eds.), *The second Glot International state-of-the-article book*. Berlin: Mouton de Gruyter. 389–429.
- Roberts, Ian (1998). Verb movement and markedness. In DeGraff, Michel (ed.), *Language creation and language change: creolization, diachrony, and development*. Cambridge, MA: MIT Press. 287–327.
- Rutherford, W. E (1982). Markedness in second language acquisition. *Language Learning* 32.1, 85–108.
- Schane, Sanford A. (1970). Phonological and morphological markedness. In Bierwisch, Manfred & Heidolph, Karl Erich (eds.), *Progress in linguistics*. The Hague: Mouton. 286–294.
- Steriade, Donca (1995). Underspecification and markedness. In Goldsmith (ed.), 114–174.
- Tiersma, Peter (1982). Local and general markedness. *Language* 58, 832–849.

- Trubetzkoy, Nikolaj (1931). Die phonologischen Systeme. *Travaux du Cercle Linguistique de Prague* 4. 96–116.
- Trubetzkoy, Nikolaus (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- Trubetzkoy, N. S. (1975). *N. S. Trubetzkoy's letters and notes*, edited by Roman Jakobson. The Hague: Mouton.
- van Riemsdijk, Henk (1978). *A case study in syntactic markedness*. Dordrecht: Foris.
- Vennemann, Theo (1988). *Preference laws for syllable structure*. Berlin: Mouton de Gruyter.
- Waugh, Linda R. (1982). Marked and unmarked: a choice between unequals. *Semiotica* 38. 299–318.
- Werner, Otmar (1989). Sprachökonomie und Natürlichkeit im Bereich der Morphologie. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42. 34–47.
- Wexler, Kenneth & Manzini, M. Rita (1987). Parameters and learnability in Binding Theory. In Roeper, Thomas & Williams, Edwin (eds.), *Parameter setting*. Dordrecht: Reidel. 41–76.
- Witkowski, Stanley R. & Brown, Cecil H. (1983). Marking-reversals and cultural importance. *Language* 59. 569–582.
- Woolford, Ellen (2001). Case patterns. In Legendre et al. (eds.), 509–543.
- Wurzel, Wolfgang U. (1984). *Flexionsmorphologie und Natürlichkeit*. Berlin: Akademie-Verlag. [English translation: Wurzel 1989.]
- Wurzel, Wolfgang U. (1987). System-dependent morphological naturalness in inflection. In Dressler et al. (eds.), 59–96.
- Wurzel, Wolfgang U. (1989). *Inflectional morphology and naturalness*. Dordrecht: Kluwer.
- Wurzel, Wolfgang U. (1994). Morphology, Natural. In Asher, R. E. (ed.), *The encyclopedia of language and linguistics* (vol. 5). Oxford: Pergamon Press. 2590–2598.
- Wurzel, Wolfgang U. (1998). On markedness. *Theoretical Linguistics* 24.1. 53–71.
- Wurzel, Wolfgang U. (2000). Inflectional system and markedness. In Lahiri, Aditi (ed.), *Analogy, levelling, markedness: principles of change in phonology and morphology*. Berlin & New York: Mouton de Gruyter. 193–214.
- Zipf, George K. (1935). *The psycho-biology of language*. Boston: Houghton Mifflin.
- Zwicky, Arnold (1978). On markedness in morphology. *Die Sprache* 24. 129–143.
- Author's address* : Max-Planck-Institut für evolutionäre Anthropologie,
Deutscher Platz 6, D-04103 Leipzig, Germany.
E-mail: haspelmath@eva.mpg.de